

1 MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPHI NREWLILLN  
 51 TGRISSYTAI GLMLGLIGQL GISLDQTRVL QNILYTASNL LLLFLGLYLS  
 101 GISSLAAKIE KIGKPIWRNL NPTLNRLLP I KSIPACLA VG ILWGWLP CGL  
 151 VYSASLYALG SGSATTGGLY MLAFALGTLP NLLAIGIFSL QLKKIMQNR  
 201 IRLCTGLSVS LWALWKLAVL WL\*

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

		10	20	30	40	50	60
10	orf103-1.pep	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRVSSYTAI					
	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLILLNTGRVSSYTAI					
		10	20	30	40	50	60
15	orf103-1.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGIISSLAAKIEKIGKPIWRNL					
	orf103ng	GLMLGLIGQLGISLDQTRVLQNILYTASNLLLLFLGLYLSGIISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
20	orf103-1.pep	NPILNRLLPKSIIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
	orf103ng	NPILNRLLPKSIIPACLA VGILWGWLP CGLVYSASLYALGSGSAATGGLYMLAFALGTLP					
		130	140	150	160	170	180
25	orf103-1.pep	NLLAIGIFSLQLKKIMQNRIRLCTGLSVSLWALWKLAVLWLX					
	orf103ng	NLLAIGIFSLQLKKIMQNRIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		

30 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 47

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

	1	ATGGA	AAACC	AAAGG	CCGCT	CCTAG	GCTTT	CGCTT	GGCAC	TTTTG	GCGGC
	51	GATGAC	GTGG	GGAAC	GCTGC	CGAT.	TCCGT	GCGGC	CAGGTA	TTGAAG	TTTG
	101	TCGATG	CGGCC	GACGCT	GGTG	TGGGT	GCGTT	TTACCG	TGGC	GGCGG	CGGTA
40	151	TTGTT	TGTTT	TGCTG	GCACT	GGGCG	GGCGG	CTGCC	aGAGC	GGCGa	GGATT
	201	TTTCT	TGGTG	CTCATT	CAGG	CTGCT	GCTGC	TCGGC	GTGC	GGGCAT	TTTCG
	251	GCAAA	CTTTG	TGCTG	ATTGC	CCAAG	GGCTG	CATTAT	ATTT	CGCCG	ACCAC
	301	GACGC	AGGTT	TTGTG	GCAGA	TTTCG	CCGTT	TACGAT	GATT	GTWGT	CGGTG
	351	TGTTG	GTGTT	TAAAG	ACCGG	ATGACT	GCCG	CTCAG	AAAAT	CGGCT	TGGTT
45	401	TTGCT	GCTTG	CCGGT	TTGCT	TATGT	ATTTT	AACGA	TAAAT	TCGGC	GAGTT
	451	GTCGG	GTTTG	GGCGC	GTATG	C.AAG	GGCGT	GTTGCT	GTGT	GCGGC	AGGCA
	501	GTATG	GCATG	GGTGT	GTAAT	GCCGT	TGGCG	AAAGC	TGCT	GTCGG	CGCAA
	551	TTCCG	GGCCG	AACAG	ATTCT	GCTGT	TGATT	TATGC	GGCAA	GTGCC	GCCGT
	601	TTTCT	TGCCG	TTTGCC	GAAAC	CGCAC	ACAT	CGGAAG	TATG	GACGG	TACGT
	651	TGGCG	TGGGT	ATGTAT	TGCG	TATTG	CTGCT	TGAAT	ACGTT	AATCG	GTATC
50	701	GGCTC	GTTCG	GCGAG	GCGTT	GAAAC	ATTGG	GAGGC	TTCCA	AAGTC	AGCGC
	751	GGTAA	CAACC	TTGCT	CCCCG	TGTTT	TACCGT	AATAA	ATACT	TTGCT	CGGGC
	801	ATTAT	GTGAT	GCCTG	AAACT	TTGCG	CGCGC	CGGA.			

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

	1	MENQR	ELLGF	RLALLA	AMTW	GTLPX	SVRQV	LKFVD	APTLV	WVRFT	VAAAV
55	51	LFVLL	ALGGR	LPKRR	DFSWC	SFRLL	LLGVA	GISAN	FVLIA	QGLHY	ISPTT
	101	TQVLW	QISPF	TMIVV	GVLFV	KDRMT	AAQKI	GLVLL	LAGLL	MYFND	KFGEL
	151	SGLGAY	XKGV	LLCAAG	SMAW	VCNAVA	QKLL	SAQFG	PQQIL	LLIYA	ASAAV
	201	FLPFA	EPAHI	GSMDG	TLAWV	CIAYC	CLNTL	IGYGS	FGEAL	KHWEA	SKVSA

251 VTTLLPVFTV INTLLGHYVM PETFAAP...

Further work revealed further partial DNA sequence <SEQ ID 401>:

```

      1 ATGGAACACC AAAGGCCGCT CCTAGGCTTC GCGTGGGCAC TTTTGGCGGC
      5 51 GATGACGTGG GGAACGCTGC CGATTGCCGT GCGGCAGGTA TTGAAGTTTG
      101 TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGCC GCGGCGGTA
      151 TTGTTGTTT TGCTGGCACT GGGCGGGCGG CTGCCGAAGC GCGGGGATT
      201 TTCTTGGTGC TCATTACAGC TGCTGCTGCT CCGCGTGGCG GGCATTTCCG
      251 CAACTTTGT GCTGATTGCC CAAGGGCTGC ATTATATTT CCGACACAG
      301 ACGCAGGTTT TGTGGCAGAT TTCGCCGTTT ACGATGATT TTGTCGGTGT
      10 351 GTTGGTGGTT AAAGACCGGA TGAAGTGGCG TCAGAAAATC GGCTTGGTTT
      401 TGCTGCTTGC CGGTTTGTCT ATGTTTTTTA ACGATAAAT CCGCGAGTTG
      451 TCGGTTTGG GCGCGTATGC GAAGGGCGTG TTGCTGTGTG CCGCAGGCAG
      501 TATGGCATGG GTGTGTTATG CCGTGGCGCA AAAGCTGCTG TCGGCGCAAT
      551 TCGGGCCGCA ACAGATTCTG CTGTTGATT ATGCGGCAAG TGCCGCGTG
      15 601 TTCCTGCCGT TTGCCGAACC GGCACACATC GGAAGTTTGG ACGGTACGTT
      651 GCGTGCGGTT TGTTTTGTCT ATTGCTGCTT GAATACGTTA ATCGGTACG
      701 GCTCGTTCCG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG
      751 TTAACAACCT TGCTCCCGCT GTTACCGTA ATAwTwwCTT TGCTCGGCA
      801 TTATGTGATG CCTGAACTT TTGCCGCGCC GGA...
  
```

20 This corresponds to the amino acid sequence <SEQ ID 402; ORF104-1>:

```

      1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV
      51 LEVLLALGGR LPKRRDFSWC SFRLLLLLGA GISANFVLI QGLHYISPTT
      101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MFFNDKFGE
      151 SGLGAYAKGV LLCAAGSMAW VCAYAVQKLL SAQFGPQIIL LLIYAASAAV
      201 FLPPFAEPAHI GSLDGTALAV CFAYCCLNTL IGYGSFGEAL KHWEASKVSA
      251 VTTLLPVFTV IXXLLGHYVM PETFAAP...
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical HI0878 protein of *H. influenzae* (accession number U32769)

ORF104 and HI0878 show 40% aa identity in 277aa overlap:

```

30 orf104 4 QRPLLGFRLALLAAMTWGTLPSXSVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 62
      Q+PLLGF AL+ AM WG+LP +++QVL ++A T+VW P
      HI0878 3 QQPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKQLPE 62

      orf104 63 --KRRDFSWCSFRLLLLLVAGISANFVLIQGLHYISPTTTQVLWQISPTTMIIVGVLVF 120
      K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F
      35 HI0878 63 LMKVRQYAW----IMLIGVIGLTSNFFLSSSLNYIEPSVAQIFIHLSSFGMLICGVLI 118

      orf104 121 KDRMTAAQKIXXXXXXXXXXXMYFNDKFGEISGLGAYXKGVLLCAAGSMAWVCNAVAQKLL 180
      K+++ OKI ++FND+F +GL Y GV+L G++ WV +AQKL+
      40 HI0878 119 KEKLGLHQKIGLFLLLIGLGLFFNDRDAFAGLNQYSTGVILGVGGALIWVAYGMAQKLM 178

      orf104 181 SAQFGPQIILLIYAASAAVFLPPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL 240
      +F QQILL++Y A F+P A+ + + + LA +C YCCCLNTLIGYGS+ EAL
      45 HI0878 179 LRKFNSQQILLMMYLGCIAIAFMADFSQVQELT-PLALICFIYCCCLNTLIGYGSYAEAL 237

      orf104 241 KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP 277
      W+ SKVS V TL+P+FT++ + + HY P FAAP
      HI0878 238 NRWDVSKVSVVITLVPLFTILFSLIAHYFSPADFAAP 274
  
```

50 Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF104 shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
55 orf104.pep MENQRPLLGFRLALLAAMTWGTLPSXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
      ||||| : |||||
      orf104a MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
      10      20      30      40      50      60

      70      80      90      100      110      120
  
```

-251-

5	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLI AQGLHYISPTTTQVLWQISPFMTIVVGVLVF
	orf104a	LPKWRDFSWCSFRLLLLGVAGISANFVLI AQGLHYISPTTTQVLWQISPFMTIVVGVLVF
		70 80 90 100 110 120
10	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL
	orf104a	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
		130 140 150 160 170 180
15	orf104.pep	SAQFGPQQILLLLIYAASAAVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL
	orf104a	SAQFGPQQILLLLIYAASAAVFLPFAELAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL
		190 200 210 220 230 240
20	orf104.pep	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP
	orf104a	KHWEASKVSAVTTLLPVFTVIFSLLLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAVG
		250 260 270 280 290 300

The complete length ORF104a nucleotide sequence <SEQ ID 403> is:

25	1	ATGGAAAACC	AAAGGCCGCT	CCTAGGCTTC	GCGTGGGCAC	TTTGGCGGCG
	51	GATGACGTGG	GGAACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
30	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGCG	GGCGGCGGTA
	151	TTGTTTGT	TGCTGGCATT	GGGCGGGCGG	CTGCCGAAGT	GGCGGGATT
35	201	TTCTTGGTGC	TCATTACAGC	TGCTGCTGCT	CGGCGTGCGC	GGCATTTCGG
	251	CAAACCTTGT	GCTGATTGCG	CAAGGGCTGC	ATTATATTTC	GCCGACCACG
40	301	ACGCAGGTTT	TGTGGCAGAT	TCGCCGTTT	ACGATGATTG	TTGTCGGTGT
	351	GTTGGTGT	AAAGACCGGA	TGACTGCCGC	TCAGAAAATC	GGCTTGGTTT
45	401	TGCTGCTTGC	CGGTTTGCTT	ATGTTTTTTA	ACGATAAATT	CGGCGAGTTG
	451	TCGGGTTTGG	GCGCGTATGC	GAAGGGCGTG	TTGCTGTGTG	CGGCAGGCAG
50	501	TATGGCATGG	GTGTGTTATG	CCGTGGCGCA	AAAGCTGCTG	TCGGCGCAAT
	551	TCGGGCCGCA	ACAGATTCTG	CTGTTGATTT	ATGCGGCAAG	TGCCGCCGTG
55	601	TTCTGCCCGT	TTGCCGAAC	GGCACACATC	GGAAGTTTGG	ACGGTACGTT
	651	GGCGTGGGTT	TGTTTTGCGT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG
60	701	GCTCGTTTCG	CGAGGGCTTG	AAACATTGGG	AGGCTTCCAA	AGTCAGCGCG
	751	GTAACAACCT	TGCTCCCGCT	GTTTACCGTA	ATATTTTCTT	TGCTCGGGCA
65	801	TTATGTGATG	CCTGATACTT	TTGCCGCGCC	GGATATGAAC	GGTTTGGGTT
	851	ATGCCGGCGC	ACTGGTCGTG	GTCGGGGGTG	CGGTACGGC	GGCGGTGGGG
	901	GACAGGCTGT	TCAAACGCCG	CTAG		

This encodes a protein having amino acid sequence <SEQ ID 404>:

45	1	MENQRPLLGF	ALALLAAMTW	GTLPIAVRQV	LKFVDAPTLV	WVRFTVAAAV
	51	LEVLLALGGR	LPKWRDFSWC	SFRLLLLGVA	GISANFVLI	AQGLHYISPTT
50	101	TQVLWQISPF	TMIVVGVLVF	KDRMTAAQKI	GLVLLLAGLL	MFFNDKFGEL
	151	SLGLAYAKGV	LLCAAGSMAW	VCYAVAQKLL	SAQFGPQQIL	LLIYAASAAV
55	201	FLPFAELAH	ISLDGTLAWV	CFAYCCLNTL	IGYGSFGEAL	KHWEASKVSA
	251	VTTLLEPVFTV	IFSLLLGHYVM	PDTFAAPDMN	GLGYAGALVV	VGGAVTAAVG
	301	DRLFRR*				

ORF104a and ORF104-1 show 98.2% identity in 277 aa overlap:

55	orf104a.pep	10 20 30 40 50 60
	orf104-1	10 20 30 40 50 60
60	orf104a.pep	70 80 90 100 110 120
	orf104-1	70 80 90 100 110 120
65	orf104a.pep	130 140 150 160 170 180
	orf104-1	130 140 150 160 170 180

10

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N. gonorrhoeae*:

25

40

45

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 407>:

50

55

60



-253-

5  
 701 GCTCGTTCGG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG  
 751 GTAACAACCT TGCTCCCCGT GTTTACCGTA ATATTTTCTT TGCTCGGGCA  
 801 TTATGTGATG CCTGATACTT TTGCCGCGCC GGATATGAAC GGTTTGGGT  
 851 ATGTCGGCGC ACTGCTCGTG GTCGGGGGTG CGGTTACGGC GGCAGTGGGG  
 901 GACAGGCCGT TCAAACGCCG CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

10  
 1 MENQRPLLGF ALALLAAMTW GTLPPIAVRQV LKFVDAPTLV WVRFTVAAAV  
 51 LEVLLALGGR LPKRRDFSWH SFRLLLLGVT GISANFVLIA QGLHYISPTT  
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLVGLL MFFNDKFGEL  
 151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV  
 201 FLPPFAEPAHI GSLDGT LAWV CFVYCCNLTL IGYGSFGEAL KHWEASKVSA  
 251 VTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYVGALVV VGGAVTAAVG  
 301 DRPFKRR\*

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

15  
 orf104-1.pep 10 20 30 40 50 60  
 MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR  
 orf104ng-1 10 20 30 40 50 60  
 MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR  
 20  
 orf104-1.pep 70 80 90 100 110 120  
 LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF  
 orf104ng-1 70 80 90 100 110 120  
 LPKRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF  
 25  
 orf104-1.pep 130 140 150 160 170 180  
 KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL  
 orf104ng-1 130 140 150 160 170 180  
 KDRMTAAQKIGLVLLLVGLLMMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL  
 30  
 orf104-1.pep 190 200 210 220 230 240  
 SAQFGPQQILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFAYCCNLTLIGYGSFGEAL  
 orf104ng-1 190 200 210 220 230 240  
 SAQFGPQQILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFVYCCNLTLIGYGSFGEAL  
 35  
 orf104-1.pep 250 260 270  
 KHWEASKVSAVTLLPVFTVIXXLLGHYVMPETFAAP  
 orf104ng-1 250 260 270 280 290 300  
 KHWEASKVSAVTLLPVFTVIFSLGHYVMPDTFAAPDMNGLGYVGALVVVGGAVTAAVG

In addition, ORF104ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

45  
 gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306  
 Score = 237 bits (598), Expect = 8e-62  
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)  
 50  
 Query: 30 QRPXXXXXXXXXXMTWGTLPPIAVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 88  
 Q+P M WG+LPIA++QVL ++A T+VW P  
 Sbjct: 3 QQPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62  
 55  
 Query: 89 --KRRDFSWHSFRLLLLGVTGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF 146  
 K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F  
 Sbjct: 63 LMKVRQYAW---IMLIGVIGLTSNFLFSSSLNYIEPSVAQIFIHLSSFGMLICGVLI 118  
 60  
 Query: 147 KDRMTAAQKIXXXXXXXXXXXMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 206  
 K+++ QKI +FFND+F +GL Y+ GV+L G++ WV Y +AQKL+  
 Sbjct: 119 KEKLGLHQKIGLFLLLIGLGLFFNDRDFAFAGLNQYSTGVILGVGALIWVAYGMAQKLM 178  
 65  
 Query: 207 SAQFGPQQILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFVYCCNLTLIGYGSFGEAL 266  
 +F QQILL++Y A F+P A+ + + L LA +CF+YCCNLTLIGYGS+ EAL  
 Sbjct: 179 LRKFNSQQILLMMYLGCIAFMPMADFSAQVQELT-PLALICFIYCCNLTLIGYGSYAEAL 237  
 Query: 267 KHWEASKVSAVTLLPVFTVIFSLGHYVMPDTFAAPDMN 306

W+ SKVS V TL+P+ET++FS + HY P FAAP++N  
 Sbjct: 238 NRWDVSKVSVVITLVPLFTILFSHIAHYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from  
 5 *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 48

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 409>:

```

10      1  ATGGTAGCTC  GTCGGGCTCA  TAACCCGAAG  GTCGTAGGTT  CGAATCCTGT
      51  .CCCGCAACC  TAATTTCAAA  CCCCTCGGTT  CAATGCCGAG  GG.GTTTTGT
     101  T.TTGCCTGT  TTCCTGTTTC  CTGTTTCCTG  CCGCCTCCGT  TTTTGGCCGG
     151  ATTTTCCTTC  CGGCCGCAAT  ATCGGAACGG  CAGACCGCCG  TCTGTTTGCG
     201  GTTGCAAATT  CAGGCAGTTT  GGCTACAATC  TTCCGCATTG  TCTCAAGAA
     251  AGCCAACCAT  GCCGACCGTC  CGTTTTACCG  AATCCGTCAG  CAAACAAGAC
     301  CTTGATGCTC  TGTTCGAGTG  GGCAAAGCA  AGTTACGGTG  CAGAAAGTTG
     351  CTGAAAACG  CTGTATCTGA  ACGGTCysCC  TTTGGGCAAC  CTGTCGCCGG
     401  AATGGGTGGA  ACGCGTsmmA  AAAGACTGGG  AGGCAGGCTG  CyCGGAGTCT
     451  TCAGACGGCA  TTTTCTGAA  TgCGGACGGc  TgGcCtGATA  TGGgCGGAcg
     501  cTTACAGCAC  CTCGCCCTCG  GTTGGCACTG  TGCGGGGCTG  TTGGACGgsT
     551  GCGCAACGA  GTGTTTCGAC  CTGACCGACG  GCGGCGGCAA  CCCCTTGTTC
     601  ACGCTCGaAc  GCGCGyTTT  mCGTCCTkTC  GGACTGCTCA  GCCGCGCCGT
     651  CCATCTCAAC  GGTCTGACCG  AATCGGACGG  CCGATGGCAT  TTCTGGATAG
     701  GCAGGCGCAG  TCCGCACAAA  GCAGTCGATC  CCAACAAACT  CGACAATACT
     751  rCCGCCGCG  GTGTTTCCGG  CGGCGAAATG  CCGTCTGAAG  CCGTGTGTCC
     801  CGAAAGCAGC  GAAGAAGCCG  GTTTGGATAA  AACCTGcTT  CCGCTCATCC
     851  GCCCGGTATC  GCAGCTGCAC  AGCCTGCGCT  CCGTCAGCCG  GGGTGTACAC
     901  AATGAAATCC  TGTATGTATT  CGATGCCGTC  CTGCCG...
```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```

30      1  MVARRAHNP  VVGSNPXPAT  XFQTPRFNAE  XVLXLPVSCF  LFPAASVFRC
     51  IFLPAAISER  QTAVCLRLQI  QAVWLQSSAL  SSRKPTMPTV  RFTESVSKQD
     101  LDALFEWAKA  SYGAESCWKT  LYLNQXPLGN  LSPWVERVX  KDWEAGCXES
     151  SDGIFLNADG  WPDMMGRLQH  LALGWHCAGL  LDGWRNECFD  LTDGGGNPLF
     201  TLERAXXRF  GLLSRAVHLN  GLTESDGRWH  FWIGRRSPHK  AVDPNKLDNT
     251  XAGGVSGGEM  PSEAVCRESS  EEAGLDKTL  PLIRPVSQLH  SLRSVSRGVH
     301  NEILYVFDAV  LP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```

40      1  ATGCCGACCG  TCCGTTTTAC  CGAATCCGTC  AGCAAACAAG  ACCTTGATGC
     51  TCTGTTTCGAG  TGGGCAAAAG  CAAGTTACGG  TGCAGAAAGT  TGCTGGAAAA
     101  CGCTGTATCT  GAACGGTCTG  CCTTTGGGCA  ACCTGTCGCC  GGAATGGGTG
     151  GAACGCGTCA  AAAAAGACTG  GGAGGCAGGC  TGCTCGGAGT  CTTACAGACG
     201  CATTTTTCTG  AATGCGGACG  GCTGGCCTGA  TATGGGCGGA  CGCTTACAGC
     251  ACCTCGCCCT  CGGTTGGCAC  TGTGCGGGGC  TGTGGACGG  CTGGCGCAAC
     301  GAGTGTTCG  ACCTGACCGA  CGGCGGCGGC  AACCCCTTGT  TCACGCTCGA
     351  ACGCGCCGCT  TTCGTCCTT  TCGGACTGCT  CAGCCGCGCC  GTCCATCTCA
     401  ACGGTCTGAC  CGAATCGGAC  GGCCGATGGC  ATTTCTGGAT  AGGCAGGCGC
     451  AGTCCGCACA  AAGCAGTCGA  TCCCAACAAA  CTCGACAATA  CTGCCGCCGG
     501  CGGTGTTTCC  GGCGCGGAAA  TGCCGTCTGA  AGCCGTGTGT  CGCGAAAGCA
     551  GCGAAGAAGC  CGGTTTGGAT  AAAACGCTGC  TTCCGCTCAT  CCGCCCGGTA
     601  TCGCAGCTGC  ACAGCCTGCG  CTCCGTCAGC  CGGGGTGTAC  ACAATGAAAT
     651  CCTGTATGTA  TTCGATGCCG  TCCTGCCCGA  AACCTTCCTG  CCTGAAAATC
     701  AGGATGGCGA  AGTGGCGGGT  TTTGAGAAAA  TGGACATCGG  CGGTCTGTTG
     751  GATGCCATGT  TGTGCGGAAA  CATGATGCAC  GACGCGCAAC  TGGTTACGCT
     801  GGACGCGTTT  TGCCGTTACG  GTCTGATTGA  TGCCGCCCAT  CCGCTGTCCG
     851  AGTGGCTGGA  CGGCATACGT  TTATAG
```

55 This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:

Computer analysis of this amino acid sequence gave the following results:

ORF105 shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) from strain A of *N*.

	60	70	80	90	100	110
orf105.pep	ISERQTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKODLDALFEWAKASYGAES					
orf105a	:					
				MPTVRFTESVSKHDLALFEWAKASYGAES		
				10	20	30
	120	130	140	150	160	170
orf105.pep	CWKTLYLNGXPLGNLSPWVERVXKDWEAGCXESSDGIFLNADGWPDMMGRLQHLALGWH					
orf105a	:					
	CWKTLYLNGLPLGNLSPWAERVKKDWEAGCESSESDGIFLNADGWPDMMGRRLQHLARIWK					
	40	50	60	70	80	90
	180	190	200	210	220	230
orf105.pep	CAGLLDGWRNECFDLTDGGGNPLFTLERAXXRPXGLLSRAVHLNGLTESDGRWHFWIGRR					
orf105a	:					
	EAGLLHGWDRDECFDLTDGGSNPLFALERAARFPGLLSRAVHLNGLVESDGRWHFWIGRR					
	100	110	120	130	140	150
	240	250	260	270	280	290
orf105.pep	SPHKAVDPNKLNTXAGGVSSGEMPSEAVCRESEEAGLDKTLPLIRPVSQQLHSLSRVS					
orf105a	:					
	SPHKAVDPDKLNTAAGGVSSGELPSETVCRESEEAGLDKTLPLIRPVSQQLHSLSRVS					
	160	170	180	190	200	210
	300	310				
orf105.pep	RGVHNEILYVFDAVLP					
orf105a						
	RGVHNEILYVFDAVLPETFLPENQDGEVAGFEKMDIGLLAAMLSGNMMHDAQLVTLDAF					
	220	230	240	250	260	270

45

55

60

1 MPTVRFESV SKHDLDAFFE WAKASYGAES CWKTLYLNGL PLGNLSPEWA  
51 ERVKKDWEAG CSESSDGIFL NADGWPMGR RLOHLARIWK EAGLLHGWRD

-256-

```

101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHLNGLVESD GRWHFWIGRR
151 SPHKAVDPAK LDNTAAGGVS SGELPSETVC RESSEEAGLD KTLPLIRPV
201 SQLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
251 AAMLSGNMMH DAQLVTDAF CRYGLIDAAH PLSEWLDGIR L*

```

5 ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

```

10 orf105a.pep      10      20      30      40      50      60
    MPTVRFTESVSKHDLDALEWAKASYGAESCWKTLYLNGPLGNLSPEWAERVKKDWEAG
    orf105-1        10      20      30      40      50      60
    MPTVRFTESVSKQDLDALEWAKASYGAESCWKTLYLNGPLGNLSPEWVERVKKDWEAG

15 orf105a.pep      70      80      90     100     110     120
    CSESSDGIIFLNADGWPDMGRRLOHLARIWKEAGLLHGWDRDECFDLTDGGSNPLFALERAA
    orf105-1        70      80      90     100     110     120
    CSESSDGIIFLNADGWPDMGRRLOHLALGWHCAGLLDGRNECFDLTDGGSNPLFLERAA

20 orf105a.pep     130     140     150     160     170     180
    FRPFGLLSRAVHLNGLVESDGRWHFWIGRRSPHKAVDPAKLDNTAAGGVSSGELPSETVC
    orf105-1        130     140     150     160     170     180
    FRPFGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPAKLDNTAAGGVSSGEMPSEAVC

25 orf105a.pep     190     200     210     220     230     240
    RESSEEAGLDKTLPLIRPVSQHLSLRPVSRGVHNEILYVDAVLPETFLPENQDGEVAG
    orf105-1        190     200     210     220     230     240
    RESSEEAGLDKTLPLIRPVSQHLSLRVSQHLNEILYVDAVLPETFLPENQDGEVAG

30 orf105a.pep     250     260     270     280     290
    FEKMDIGGLLAAMLSGNMMHDAQLVTDAFCRYGLIDAAHPLSEWLDGIRLX
    orf105-1        250     260     270     280     290
    FEKMDIGGLLDAMLSGNMMHDAQLVTDAFCRYGLIDAAHPLSEWLDGIRLX

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N. gonorrhoeae*:

```

40 orf105.pep      MVARRAHNPKVVGSNPXATXFTQTPRFNAEXVLXLPVSCFLFPAASVFCRIFLPAISER 60
    orf105ng        MVARRAHNPKVVGSNPAPATKYQTPRFNAEGVLF-----FLFPAASVFCRIFLPAISER 55

45 orf105.pep      QTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAESCWKT 120
    orf105ng        QAAVCLRLQIQAVWLQSSALCSRKPAMPTVRFTESVSKQDLDALEWAKASYGAESCWKT 115

50 orf105.pep      LYLNGLPLGNLSPEWVERVXKDWEAGCXESSDGIIFLNADGWPDMGRRLOHLALGWHCAGL 180
    orf105ng        LYLNRLPLGNLSPEWAERIKKWEAGCSESSNGIFLNADGWPDMGRRLOHLARTWNKAGL 175

55 orf105.pep      LDGWRNECFDLTDGGSNPLFTLERAXRXPXGLLSRAVHLNGLTESDGRWHFWIGRRSPHK 240
    orf105ng        LHGWRNECFDLTDGGSNPLFTLERAAFRPFGLLIRAVHLNGLVESNGRWHFWIGRRSPHK 235

60 orf105.pep      AVDPNKLNTXAGGVSGGEMPSEAVCRSEEEAGLDKTLPLIRPVSQHLSLRVSQHLNEILYVDAVLP 300
    orf105ng        AVDPGKLDNIAGGVSGGEMPSEAVCRSEEEAGLDKTLFPLIRPVSRLHSLRPVSRGVH 295

    orf105.pep      NEILYVDAVLP 312
    orf105ng        NEILYVDAVLPETFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLVTDAFYRYG 355

```

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:

1	MVARAHNP	KVGSNPAPAT	KYQTPRFNAE	GVLFFLFPA	SVFCRIFLPA	
51	ATSERQA	AVLR	LQIQAVWL	QSSSLCSRKP	AMPTVRFTES	VSKQDLDALF
101	ERAKASY	GAE	CRKWTLYLNR	LPLGNLSPEW	AERIKKDWEA	GCSESSNGIF
151	LNADGWP	DMG	GRQLHLARTW	NKAGLLHGWR	NECFDLDG	GNPLFTLERA
201	AFRPFGL	LIR	AVHLNGLVES	NGRWHFWIGR	RSPHKAVDPG	KLDNIAGGGV
251	SGGEMP	SEAV	CRESSEEGAL	DKTFLPLIRP	VSRLHSLRPV	SRGVHNEILY
301	VFDAVL	PETF	LPLENGDGEA	GFEKMDIGGL	LDAMLSKNMM	HDAQLVTLDA
351	FYRYGL	DA	HLSEWLDGI	RL*		

10

15

20

25

30

1	MPTVRFTESV	SKQDLDALE	RAKASYGAES	CWKTLYLNRL	PLGNLSPEWA
51	ERIKKDWAG	CSESSDGIFL	NADGWPMGSR	RLQHLARTWN	KAGLLHGWRN
101	ECFDLTDGG	NLFTFLERA	FRPGLLSRA	VHLNGLVESN	GRWHFWIGRN
151	SPHKAVDPGK	LDNIAGGGVS	GGEMPSEAVC	RESSEEAAGL	KTLPFLIRPV
201	SRLHSLRPVS	RQVHNILYV	FDVFLPETFL	PENQDQGEAV	FEKMDIGGLL
251	DAMLSKNMHH	DAQVLTDAF	YRYGLIDAAH	PLSEWLDGIR	I.*

35 ORG105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

40

```

              10      20      30      40      50      60
orf105-1.pep  MPTVRFTESVSKQDLDALEWAKASYGAESCWKTLYNLGLPLGNLSPWEVVERVKDWEAG
               |||||
orf105ng-1    MPTVRFTESVSKQDLDALEWAKASYGAESCWKTLYNLRLPLGNLSPWEAERIKDWEAG

```

45

```

              70              80              90              100              110              120
orf105-1.pep  CSESSDGI FLNADGW PDMGGRLQHLALGWHCAGLLDGWRNECFDLTDGGGNNPLFTLERA
              |||||
orf105ng-1    CSESSDGI FLNADGW PDMGGRLQHLARTWNKAGLLHGWRNECFDLTDGGGNNPLFTLERA
              70              80              90              100              110              120

```

50

```

              130      140      150      160      170      180
orf105-1.pep  FRPFGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPNKLNTAAGGVSGGEMPSEAVC
              |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orf105ng-1    FRPFGLLSRAVHLNGLVESNGRWHFWIGRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVC
              130      140      150      160      170      180

```

55

```

              190      200      210      220      230      240
orf105-1.pep  RESSEEAGLDKTLPLIRPVSQLHSLRSVSRGVHNEILYVFDAVLPETFLPENQDGEVAG
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf105ng-1    RESSEEAGLDKTLFPLIRPVSRSLHSLRPVSRGVHNEILYVFDAVLPETFLPENQDGEVAG
              190      200      210      220      230      240

```

60

```

                250      260      270      280      290
orf105-1.pep  FEKMDIGGLLDAMLSGNMMHDAQLVLTDAFCRYGLIDAAHPLSEWLDGIRLX
               |||||
orf105ng-1    FEKMDIGGLLDAMLSKNMMHDAQLVLTDAFYRYGLIDAAHPLSEWLDGIRLX
               250      260      270      280      290

```

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

```

5  sp|P41888|TNR3_SCHPO THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
   >gi|1076928|pir|S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
   (Schizosaccharomyces pombe) >gi|666111 (X84417) thiamin pyrophosphokinase
   [Schizosaccharomyces pombe] >gi|2330852|gnl|PID|e334056 (Z98533) thiamin
   pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
   Score = 105 bits (259), Expect = 4e-22
   Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)

10  Query: 268 NKAGLLHGWRNECFDLTDGGGNPLFTLERAARFPFGLLSRAVHLNGLVESNGRW--HFWI 441
   N G+ WRNE + + P+ +ER F FG LS VH + + W+
   Sbjct: 96 NTFGIADQWRNELYTVYGKSKKPVLAVERGGFWLFGFLSTGVHCTMYIPATKEHPLRIWV 155

15  Query: 442 GRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVCRESSEEAGLDKTLFPLIRPVSRLHSLR 621
   RRSP K P LDN GG++ G+ + +E SEEA LD + LI P + ++
   Sbjct: 156 PRRSPTKQTWPNYLDNSVAGGIAHGDSVIGTMIKEFSEANLDVSSMNL-PCGTVSYIK 214

   Query: 622 PVSRG-VHNEILYVFDAVLPEFLENQDGEVAGFEKMDIGLLDAML SKNMHDAQLVT 798
   R + E+ YVFD + + +P DGEVAGF + + +L + K+ + LV
20  Sbjct: 215 MEKRHWIQPELQYVFDDLVPINDGEVAGFSLPLNQVLHELELKSFKPNCALVL 274

   Query: 799 LDAFYRYGLIDAAHP 843
   LD R+G+I HP
   Sbjct: 275 LDFLIRHGIITPQHP 289

```

25 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 30 419>:

```

1  ATGAATAGAC CCAAGCAACC CTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
51  CCAAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
101 CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
35 151 TTGATATTG GTAACTATAC GCGAAAGACA ACAGTGGAGG GACAAATTTT
201 ACCTGCATCG GCGTAATCA GGGTGTATGC ACCGgATACG rGkACAATTA
251 CAGCGAAATT CGTGAAGAT GGmsAAAAGG TTAAGGCTGG CGACAAGCTA
301 TTTGCGCTTT CGACCTCACG TTTCGGCGCA GGAGGTAGCG TGCAGCAGCA
351 GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAAC TGG
40 401 GTCGTCTGAA GCTGATACAC GGAATGAAA CGCGCAGCcT TAAAGCAACT
451 GTCGAACGTT TGGAAAACCA GGAAC TCCAT ATTTCCGCAAC AGATAGACGG
501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
551 TCCTATCCGC .CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 420; ORF107>:

```

45 1  MNRPKQPFFR PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
   51  LIFGNYTRKT TVEGQILPAS GVIRVYAPDT XTITAKFVED GXKVKAGDKL
101  FALSTSRFGA GGSVQQQLKT EAVLKKT LAE QELGR LKLIH GNETRSLKAT
151  VERLENQELH ISQQIDGQKR RIRLAE EMLQ KYRFLSXQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.meningitidis*:

10 20 30 40 50 60

-259-

	orf107.pep	MNRPKQPFFRPEVAVARQTS	LTGKVILTRPLSFSLWTTFA	SISALLIILFLIFGNYTRKT	
	orf107a	MNRPKQPFFRPEVAVARQTS	LTGKVILTRPLSFSLWTTFA	SISALLIILFLIFGNYTRKT	
5		10 20 30 40 50 60			
	orf107.pep	TVEGQILPASGVIRVYAPD	XTITAKFVEDGXKVKAGDKL	FALSTSRFGAGGSVQQQLKT	
	orf107a	TVEGQILPASGVIRVYAPD	TGTITAKFXEDGKVKAGDKL	FALSTSRFGAGDSVQQQLKT	
10		70 80 90 100 110 120			
	orf107.pep	EAVLKKTAEQELGRLKLIH	GNETRSLKATVERLENQELH	ISQQIDGQKRRIRLAEEMLQ	
	orf107a	EAVLKKTAEQELGRLKLIH	GNETRSLKATVERLENQELH	ISQQIDGQKRRIRLAEEMLQ	
15		130 140 150 160 170 180			
	orf107.pep	KYRFLSXQX			
	orf107a	KYRFLSANDAVPKQEMMNV	KAELEQKAKLDAYRREEVGL	LQEIRTQNLTLXSLPQAA	
20		189			
		190 200 210 220 230			

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

	1	ATGAATAGAC	CCAAGCAACC	NTTCTTCCGT	CCCGAAGTCG	CCGTTGCCCG
25	51	CCAAACCAGC	CTGACGGGTA	AAGTGATTCT	GACACGACCG	TTGTCATTTT
	101	CCCTATGGAC	GACATTTGCA	TCGATATCTG	CGTTATTGAT	TATCCTGTTT
	151	TTGATATTG	GTAACATATC	GCGAAAGACA	ACAGTGGAGG	GACAAATTTT
	201	ACCTGCATCG	GGCGTAATCA	GGGTGTATGC	ACCGGATACG	GGGACAATTA
30	251	CNGCGAAATT	CNTGGAAGAT	GGAGAAAAGG	TTAAGGCTGG	CGACAAGCTA
	301	TTTGCCTTT	CGACCTCACG	TTTCGGCGCA	GGAGATAGCG	TGCAGCAGCA
	351	GTTGAAAACG	GAGGCAGTTT	TGAAGAAAAC	GTTGGCAGAA	CAGGAAGTGG
	401	GTCGTCTGAA	GCTGATACAC	GGGAATGAAA	CGCGCAGCCT	TAAAGCAACT
	451	GTCGAACGTT	TGGAAAACCA	GGAATCCAT	ATTTCCGCAAC	AGATAGACGG
35	501	TCAGAAAAGG	CGCATTAGAC	TTGCGGAAGA	AATGTTGCAG	AAATATCGTT
	551	TCCTATCCGC	CAATGATGCA	GTGCCAAAAC	AAGAAATGAT	GAATGTCAAG
	601	GCAGAGCTTT	TAGAGCAGAA	AGCCAACTT	GATGCCTACC	GCCGAGAAGA
	651	AGTCGGGCTG	CTTCAGGAAA	TCCGCACGCA	GAATCTGACA	TTGGNNAGCC
	701	TCCCCCAAGC	GGCATGA			

This encodes a protein having amino acid sequence <SEQ ID 422>:

40	1	MNRPKQPFFR	PEVAVARQTS	LTGKVILTRP	LSFSLWTTFA	SISALLIILF
	51	LIFGNYTRKT	TVEGQILPAS	GVIRVYAPDT	GTITAKFXED	GEKVKAGDKL
	101	FALSTSRFGA	GDSVQQQLKT	EAVLKKTAE	QELGRLKLIH	GNETRSLKAT
	151	VERLENQELH	ISQQIDGQKR	RIRLAEEMLQ	KYRFLSANDA	VPKQEMMNVK
45	201	AELLEQKAKL	DAYRREEVGL	LQEIRTQNL	LXSLPQAA*	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N.gonorrhoeae*:

50	orf107.pep	MNRPKQPFFRPEVAVARQTS	LTGKVILTRPLSFSLWTTFA	SISALLIILFLIFGNYTRKT	60
	orf107ng	MNRPKQPFFRPEVAIARQTS	LTGKVILTRPLSFSLWTTFA	SISALLIILFLIFGNYTRKT	60
	orf107.pep	TVEGQILPASGVIRVYAPD	XTITAKFVEDGXKVKAGDKL	FALSTSRFGAGGSVQQQLKT	120
55	orf107ng	TMEGQILPASGVIRVYAPD	TGTITAKFVEDGKVKAGDKL	FALSTSRFGAGGSVQQQLKT	120
	orf107.pep	EAVLKKTAEQELGRLKLIH	GNETRSLKATVERLENQELH	ISQQIDGQKRRIRLAEEMLQ	180
60	orf107ng	EAVLKKTAEQELGRLKLIH	ENETRSLKATVERLENQKLH	ISQQIDGQKRRIRLAEEMLR	180
	orf107.pep	KYRFLSXQ	188		
	orf107ng	KYRFLSAQ	188		

The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

```

1  MNRPKQFFFR PEVAIARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
51  LIFGNYTRKT TMEGQILPAS GVIRVYAPDT GTITAKFVED GEKVKAGDKL
101 FALSTSRFGA GGSVQQQLKT EAVLKKTLE QELGRLKLIH ENETRSLKAT
151 VERLENQKLH ISQQIDGQKR RIRLAEEMLR KYRFLSAQ*

```

Based on the presence of a putative ransmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 10 Example 50

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

```

1  ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGCC
51  GTGCGGCAAA TCCGTAAATA CGGCGGTACA GCCGCAAAAC GCGGTACAAA
151 GCGCGCCGAA ACCGGTTTTC AAAGTCATAT ATATCGACAA TACGGCGATT
201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GCGCACTTGG AAGCCGTCAG CGGCAAATGT
301 ATGGAACCG ATGATAAGGA CAGTCCGGCA GGTGCGGCAG AAAACGCGGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTTCCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

```

1  MLNTFFAVLG GCLLXLPCGK SVNTAVQPQN AVQSAPKPVF KVIYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDSPA GWAENGVCHT LFAKLVGNIA EDGGKLT DYL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFFRRRH Y*

```

Further work revealed the following DNA sequence <SEQ ID 427>:

```

1  ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTGCCTGC TGCTTGCCGC
51  CTGCGGCAAA TCCGAAAATA CGGCGGAACA GCCGCAAAAC GCGGTACAAA
101 GCGCGCCGAA ACCGGTTTTC AAAGTCAAAT ATATCGACAA TACGGCGATT
151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GCGCACTTGG AAGCCGTCAG CGGCAAATGT
301 ATGGAACCG ATGATAAGGA CAGTCCGGCA GGTGCGGCAG AAAACGCGGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTTCCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

```

1  MLKTSFAVLG GCLLLAACGK SENTAEQPQN AVQSAPKPVF KVKYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDSPA GWAENGVCHT LFAKLVGNIA EDGGKLT DYL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFFRRRH Y*

```

Computer analysis of this amino acid sequence gave the following results:



Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) from *N.gonorrhoeae*:

```

5      orf108.pep  MLNTFFAVLGGCLLXLPCGKSVNTAVQPQNAVQSAPKPVFKVIYIDNTAIAGLDLGQSSE  60
      orf108ng    MLKIPFAVLGGCLLLAACGKSENTAEQPQNAAQSAPKPVFKVKYIDNTAIAGLALGQSSE  60

10     orf108.pep  GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      orf108ng    GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

15     orf108.pep  LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH  181
      orf108ng    LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH  181

```

ORF108-1 shows 92.3% identity with ORF108ng over the same 181 aa overlap:

```

20     orf108-1.pep MLKTSFAVLGGCLLLAACGKSENTAEQPQNAVQSAPKPVFKVKYIDNTAIAGLDLGQSSE  60
      orf108ng-1    MLKIPFAVLGGCLLLAACGKSENTAEQPQNAAQSAPKPVFKVKYIDNTAIAGLALGQSSE  60

25     orf108-1.pep GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      orf108ng-1    GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

30     orf108-1.pep LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH  181
      orf108ng-1    LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFYFRRRH  181

```

The complete length ORF108ng nucleotide sequence <SEQ ID 429> is:

```

30      1  ATGCTGAAAa  tacctTTTGC  CGTGtgggc  ggCtgcctGC  TGCTTGCCGC
      51  CTGCGGCAAA  TCCGAAAATa  cggcggaACA  GCCGCAAAAT  gcggCACAAA
      101  GCGCGCCGAA  ACCGGTTTTC  AAAGTCAAAT  ACATCGACAA  TACGGCGATT
      151  GCCGGTTTGG  CTTTGGGACA  AAGTAGCGAA  GGCAAAACCA  acgacgGCAA
      201  AAAACAATC  AGTTATccgA  TTAAAGGCTT  GCCGGAACAA  Aacgccgtcc
      251  gGCTGACCGG  AAAGCATCCC  AACGACTTGG  AagccgtcgT  CGGCAAATGT
35      301  ATGGAACCG  ACGGAAAGGA  CGCGCCTTCG  GGCTGGGCGG  AAAACGGCGT
      351  GTGCCATACC  TTGTTTGCCA  AACTGGTGGG  CAATATCGCC  GAAGACGGCG
      401  GCAAACTGAC  TGATTACCTG  ATTTGCGATT  CCGCCCTGCA  ACCCTATCAG
      451  GCAGGCAAAA  GCGGCTATGC  CGCCGTGCAG  AACGGACGCT  ATGTGCTGGA
      501  AATCGACAGC  GagggGGCGT  TTTATttccg  ccgccgccat  tattgA

```

40 This encodes a protein having amino acid sequence <SEQ ID 430>:

```

      1  MLKIPFAVLG  GCLLLAACGK  SENTAEQPQN  AAQSAPKPVF  KVKYIDNTAI
      51  AGLALGQSSSE  GKTNDGKKQI  SYPIKGLPEQ  NAVRLTGKHP  NDLEAVVGKC
      101  METDGKDAPS  GWAENGVCHT  LFAKLVGNI  AEDGGKLTDL  ISHSALQPYQ
      151  AGKSGYAAVQ  NGRYVLEIDS  EGAFYFRRRH  Y*

```

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 50 Example 51

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

```

      1 ATGGAAGATT TATATATAAT ACTCGCTTTG GTTTGGTTG CGATGATTGC
    51 CGGATTATATC GATgcatTg cGggCGGGG TGGTTTGATT ACGCTGCCCC
   101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
   151 CTGCAAgCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA
   201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
   251 TAGGCGGCGT GGcCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
   301 CTgCTgGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT
   351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
   401 TTTTCTGTT CGGGCTGACG GTCGC .ACCG CTTTGGGTT TTTACGACGG
   451 TGTGTTGCGA CCGGGTGTCG GCTCGTTTTT TCTGATTGCC TTTATTGTTT
   501 TGCTCGGCTG CAAgCTGTG AACGCGATGT CTTACACCAA ATTGGCGAAC
   551 GTTGCCTGCA ATCTTGGTTC GCTATCGGTA TTCTGCTGC ACGGTTCGAT
   601 TATTTTCCCG ATTGCGGCAA CGaTGGCGGT CGGTGCGTTT GTCGgtGCGA
   651 ATTTAgGTGC GAGATTTGCC GTaCgctTCG GTTCGAAGCT GATTAA

```

15 This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

```

      1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
    51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
   101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLEGLT VXTAFGLRR
   151 CVRTGCRLVF SDCLYCFARL QAVRDVLHQ IGERCLQSWF AIGIPAARFD
   201 YFPDCGNDGG RCVCRCEFRG EICRTLRFEA D*

```

Further work revealed the following DNA sequence <SEQ ID 433>:

```

      1 ATGGAAGATT TATATATAAT ACTCGCTTTG GTTTGGTTG CGATGATTGC
    51 CGGATTATATC GATGCGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCCC
   101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
   151 CTGCAAGCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA
   201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
   251 TAGGCGGCGT GGCCCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
   301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT
   351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
   401 TTTTCTGTT CGGGCTGACG GTCGCACCGC TTTTGGGTTT TTACGACGGT
   451 GTGTTTCGGAC CGGGTGTGCG CTCGTTTTTT CTGATTGCCT TTATTGTTT
   501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAAA TTGGCGAACG
   551 TTGCCCTGCAA TCTTGGTTCG CTATCGGTAT TCCTGCTGCA CGGTTCGATT
   601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
   651 TTTAGGTGCG AGATTTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
   701 TGCTGATTGT CATCAGCATT TCATGGCTG TGAAATTGTT GATAGACGAG
   751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 434; ORF109-1>:

```

      1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
    51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
   101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLEGLT VAPLLGFYDC
   151 VFGPGVGSEF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS I
   201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLVISI SMAVKLLIDE
   251 RNPLYQMIVS MF*

```

45 Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
50 orf109.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
    orf109a  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      10      20      30      40      50      60
55 orf109.pep TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
    orf109a  TVSFARKGLIDWKKGLPIAAASFAGGVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
      70      80      90     100     110     120

```

```

      130      140      150      160      170      180
orf109.pep KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ
5 orf109a   KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
      130      140      150      160      170      180

```

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

```

      1 ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTCGGTTG CGATGATTGC
      51 CCGATTTATC GATGCGATTG CGGGTGGGGG TGGTTTGATT ACGCTGCCTG
10 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
      151 CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
      201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCGGCA GCATCGTTTG
      251 CAGGCGGCGT GGTTCGGTCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
      301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
15 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
      401 TTTTCTGTGT CCGTCTGACG GTTGCAACCAC TTTTGGGTTT TTACGACGGT
      451 GTGTTCCGAC CGGGTGTCGG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
      501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAACC
      551 TTGCCTGCAA TCTTGGTTCG CTATCGGTAT TCCTGCTGCA CGGTTCGATT
20 601 ATTTTCCCGA TTGCAGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
      651 TTTAGGTGCG AGATTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
      701 TGTGATTGT CATCAGCATT TCGATGGCTG TGAATTTGTT GATAGACGAG
      751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 436>:

```

25      1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
      51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
10 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
      151 VFGPGVGSFF LIAFIVLLGC KLLNAMSYTK LANVACNLGS LSVFLLHGS
20 201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE
30 251 RNPLYQMIVS MF*

```

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

```

      10      20      30      40      50      60
orf109a.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
35 orf109-1   MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      10      20      30      40      50      60

      70      80      90      100     110     120
orf109a.pep TVSFARKGLIDWKKGLPIAAASFAGGVVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
40 orf109-1   TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
      70      80      90      100     110     120

      130     140     150     160     170     180
orf109a.pep KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
45 orf109-1   KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
      130     140     150     160     170     180

      190     200     210     220     230     240
orf109a.pep LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
50 orf109-1   LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
      190     200     210     220     230     240

      250     260
orf109a.pep SMAVKLLIDERNPLYQMIVSMFX
55 orf109-1   SMAVKLLIDERNPLYQMIVSMFX
      250     260
60

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N.*

*gonorrhoeae*:

```

5      orf109.pep  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60
      orf109ng    MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60

      orf109.pep  TVSFARKGLIDWKKGLPIAAASFVGGVAGALS SVSLVSKDILLAVVPVLLIFVALYFVFSP  120
      orf109ng    TVSFARKGLIDWKKGLPIAAASFAGGVGALS SVSLVSKDILLAVVPVLLIFVALYFVFSP  120

      orf109.pep  KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ  180
      orf109ng    KLDGSKEGKARMSFFLFGLTVATAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ  180

15     orf109.pep  IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRTLRFEAD  231
      orf109ng    IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRPLRFEAD  231

```

An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino

acid sequence <SEQ ID 438>:

```

1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VATAFGFLRR
151 CVRTGCRLVF SDCLYCFARL QAVERDVLHQ IGERCLQSWF AIGIPAARFD
25 201 YFPDCGNDGG RCVCRCEFR C EICRPLRFEA D*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 439>:

```

1  ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTCGGTTG CGATGATCGC
51 CGGATTTATC GATGCGATTG CGGGCGGGGG TGGTTTGATT ACGCTGCCTG
101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAG
151 CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TGCACGCAA
201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
251 CAGCGGCGGT GGTGCGTGCA TTATCGGTCA GCTTGCTTTC CAAAGATATT
301 TTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGCG TGTATTTTGT
351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401 TTTTCTTATT CGGGCTGACG GTTGCAACGC TTTTGGGTTT TTACGACGGT
451 GTGTTTCGGAC CGGGTGTCGG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAAA TTGGCGAAGC
551 TTGCTTGCAA TCTTGCTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
601 ATTTTCCCGA TTGTGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
40 651 TTTAGGTGCG AGATTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
701 TGTGATTGT CATCAGCAT TCGATGGCTG TGAATTTGTT GATAGACGAG
751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

```

1  MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
45 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
151 VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS I
201 IFPIVATMAV GAFVGANLGA RFAVREGSKL IKPLLVISI SMAVKLLIDE
251 RNPLYQMIVS MF*

```

ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

```

      10      20      30      40      50      60
orf109ng-1.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      |||||
orf109-1        MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      10      20      30      40      50      60

      70      80      90     100     110     120
orf109ng-1.pep TVSFARKGLIDWKKGLPIAAASFAGGVGALS SVSLVSKDILLAVVPVLLIFVALYFVFSP

```

-265-

```

|||||:|||||
orf109-1 TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
          70      80      90      100     110     120
5
          130     140     150     160     170     180
orf109ng-1.pep KLDGSKEGKARMSFFLEGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
|||||
orf109-1 KLDGSKEGKARMSFFLEGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
          130     140     150     160     170     180
10
          190     200     210     220     230     240
orf109ng-1.pep LANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGGANLGARFAVRFGSKLIKPLLIVISI
|||||:|||||
orf109-1 LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGGANLGARFAVRFGSKLIKPLLIVISI
          190     200     210     220     230     240
15
          250     260
orf109ng-1.pep SMAVKLLIDERNPLYQMIVSMFX
|||||
20
orf109-1 SMAVKLLIDERNPLYQMIVSMFX
          250     260

```

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

```

sp|P29942|YCB9 PSEDE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3'REGION (ORF9)
>gi|94984|pir||I38164 hypothetical protein 9 - Pseudomonas sp >gi|551929
25 (M62866) ORF9 [Pseudomonas denitrificans] Length = 261
   Score = 175 bits (439), Expect = 3e-43
   Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)

Query: 41 PPVSAIATNKLQXXXXXXXXXXXXXRRKGLIDWKKGLPIXXXXXXXXXXXXXXXXXXXXKDI 100
30 PP+ + TNKLQ R+G ++ K+ LP+ D+
Sbjct: 43 PPLQTLGTNKLQGLFGSGSATLSYARRGHVNLKEQLPMALMSAAGAVLGALLATIVPGDV 102

Query: 101 LLAVVPVLLIFVALYFVFSPKLDGSKEGKARMSFFLEGLTVAPLLGFYDGVFGPGVGSFF 160
35 L A++P LLI +ALYF P + G + +R++ F+F LT+ PL+GFYDGVFGPG GSFF
Sbjct: 103 LKAILPFLIIAIALYFGLKPNM-GDVDQHSRVTPFVFTLTIVPLIGFYDGVFGPGTGSFF 161

Query: 161 LIAFIVLLGCKLLNAMS YTKLANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGGANLGA 220
++ F+ L G +L A ++TK N N+G+ VFL G++++ + M +G F+GA +G+
40 Sbjct: 162 MLGFVTLAGFGVLKATAHTKFLNFGSNVGAFGVFLFFGAVLWKVGLLMGLGQFLCAQVGS 221

Query: 221 RFAVRFGSKLIKPLLIVISISMAVKLLIDERNPL 254
R+A+ G+K+IKPLL+++SI++A++LL D +PL
Sbjct: 222 RYAMAKGAKIIKPLLIVIVISIALAIRLLADPTPL 255

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 52

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

```

50      1 ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAAACCAG
      51 CCGCAGACGG ATTATTTGGT CAAATTCGGA TCGTTTTGGG CGAG.ATTTT
      101 TGGTTTTCTG GGA CTGTATG ACGTCTATGC TTCGGCATGG TTTGTCGTTA
      151 TCATGATGTT TTTGGTGGTT TCTACCA GTT TGTGCCTGAT TCGCAATGTG
      201 CCGCCGTTCT GCGCGGAAAT GAAGTCTTTT CGGGAAAAGG TTAAAGAAAA
55      251 ATCTCTGGCG GCGATGCGCC ATTCTTCGCT GTTGGATGTA AAAATTGCGC
      301 CCGAGGTTGC CAAACGTTAT CTGGAAGTAC AAGGTTTCA GGGGAAAACC
      351 ATTAACCGTG AAGACGGGTC GGTTCGTATT GCCGCCAAA AAGGCACAAT
      401 GAACAAATGG GGCTATATCT TTGCCCATGT TGCTTTGATT GTCATTTGCC
      451 TGGGCGGTT GATAGACAGT AACCTGCTGT TGAAACTGGG TATGCTGACC
60      501 GGTCGGATTG TTCCGGACAA TCAGGCGGTT TATGCCAAGG ATTC.AAGC

```

551 CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTTAGGG GCAACGTCAA  
601 TATTTCCG.A GGGGCAGAgT GCGGATGTGG TTTTCCTGA

This corresponds to the amino acid sequence <SEQ ID 442; ORF110>:

5           1   ..LLGIASVIGT LLQONQPQTD YLVKFGSFWA XIFGFLGLYD VYASAWFVVI  
          51   MMFLVVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKIAP  
         101   EVAKRYLEVQ GFQGKTINRE DGSVLIAAKK GTMNKWGYIF AHVALIVICL  
         151   GGLIDSNLLL KLGMLTGRIF RTIRREMPRI XKPESXFGCV QSLI\*GQRQY  
         201   FXRGRVRMWF S\*

Computer analysis of this amino acid sequence gave the following results:

#### 10   Homology with ORF88a from *N.meningitidis* (strain A)

ORF110 shows 91.5% identity over a 188aa overlap with ORF88a from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf88a.pep	MSKSRRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA					
15	orf110			LLGIASVIGTLLQONQPQTDYLVKFGSFWA			
				10	20	30	
		70	80	90	100	110	120
20	orf88a.pep	QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH					
	orf110	XIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH					
		40	50	60	70	80	90
25	orf88a.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL					
	orf110	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL					
		100	110	120	130	140	150
30	orf88a.pep	GGLIDSNLLLKLGLMTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF					
	orf110	GGLIDSNLLLKLGLMTGRIFRTIRREMPRIKXKPESXFGCVQSLIXGQRQYFXRGRVRMWF					
35		160	170	180	190	200	210
	orf88a.pep	LNADNGILVQDLPEVVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLETRIVNHPIT					
	orf110	SX					

40   However, ORF88 and ORF110 do not align, because they represent two different fragments of the same protein.

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) from *N. gonorrhoeae*:

45	orf110.pep	LLGIASVIGTLLQONQPQTDYLVKFGSFWA	30
	orf110ng	MSKSRIPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGPFWT	60
50	orf110.pep	XIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	90
	orf110ng	RIFDFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf110.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKWGYIFAHVALIVICL	150
55	orf110ng	SSLLDVKIAPEVAKRYLEVGRGQKTVSREDGSVLIAAKKGTMNKWGYIXAHVALIVICL	180

-267-

```

orff110.pep  GGLIDSNLLLLKLGMLTGRIFRTIRRFMPRIKKPESXFGCVQSLIXGQRQYFXRGRVRMWF  210
|  ||: ||||| ||||: | |||: || |||| ||||: | |||| ||||| ||: |||||
orff110ng    GRLINXNLLLLKLGMLAGSIFRNRRVMPRIKSPESIWGGVQSLIKGQRQYFQRGKVRMWF  240

5  orff110.pep  S  211
   |
   orff110ng    S  241

```

The complete length ORF110ng nucleotide sequence <SEQ ID 443> is predicted to encode a protein having amino acid sequence <SEQ ID 444>:

```

10      1  MSKSRIPTL  LSRPWFAPFS  SMRFAVALLS  LLGIASVIGT  VLQONQPQTD
      51  YLVKFGPFWT  RIFDFLGLYD  VYASAWFVVI  MMFLVVSTSL  CLIRNVPPFW
     101  REMKSFREKV  KEKSLAAMRH  SSLLDVKIAP  EVAKRYLEVR  GFQGKTVSRE
     151  DGSVLIAAKK  GTMNKWKGYI  AHVALIVICL  GRLINXNLLL  KLGMLAGSIF
     201  RNNRRVMPRI  SKPESIWGGV  QSLIKGQRQY  FQRGKVRMWF  S*

```

Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 53

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 445>:

```

20      1  ATGCCGTCTG  AAACACGCCT  GCCGAAC TTT  ATCCGCGTCT  TGATATTTGC
      51  CCTGGGTTC  ATCTTCTCTGA  ACGCCTGTTC  GGAACAAACC  GCGCAAACCG
     101  TTACCCTGCA  AGGCGAAACG  ATGGGCACGA  CCTATACCGT  CAAATACCTT
     151  TCAAATAATC  GGGACAAACT  CCCCTCACCT  GCCGAAATAC  AAAAACGCAT
     201  CGATGACGCG  CTTAAAGAAG  TCAACCGGCA  GATGTCCACC  TATCAGCCCG
     251  ACTCCGAAAT  CAGCCGGTTC  AACCAACACA  CAGCCGGCAA  GCCCCTCCGC
     301  ATTTCAAGCG  ACTTCGCACA  CGTTACTGCC  GAAGCCGTCC  GCCTGAACCG
     351  CCTGACACAC  GGC GCGCTGG  ACGTAACCGT  CGGCCCTTG  GTCAACCTTT
     401  GGGGATTTCG  CCCCACAAA  TCCGTTACCC  GTGAACCGTC  GCCGGAACAA
     451  ATCAAACAGG  CGGCATCTTA  TACGGGCATA  GACAAAATCA  TTTTGAAACA
     501  AGGCAAAGAT  TACGCTTCCT  TGAGCAAAAC  CCACCCCAAG  GCCTATTTGG
     551  ATTTATCTTC  GATTGCCAAA  GGCTTCGGCG  TTGATAAAGT  TGC GGGCGAA
     601  CTGGAAAAAT  ACGGCATTCA  AAATTATCTG  GTCGAAATCG  CGGCGAGTT
     651  GCACGGCAAA  GGCAAAAACG  CGCGCGGCGA  ACCGTGGCGC  ATCGGTATCG
     701  AGCAGCCCAA  TATCGTCCAA  GGCGGCAATA  CGCAGATTAT  CGTCCCGCTG
     751  AACAAACGTT  CGCTTGCCAC  TTCCGCGCAT  TACCGTATTT  TCCACGTCGA
     801  TAAAAACGGC  AAACGCCTCT  CCCATATCAT  CAACCCGAAC  AACAAACGAC
     851  CCATCAGCCA  CAACCTCGCC  TCCATCAGCG  TGGTCGCAGA  CAGTGCGATG
     901  ACGGCGGACG  GCTTGTCAC  AGGATTATTC  GTATTGGGCG  AAACCGAAGC
     951  CTTAAAGCTG  GCAGAGCGCG  AAAAATCGC  TGTTTTCTCT  ATTGTCAGGG
    1001  ATAAAGCGG  CTACCGCACC  GCCATGTCTT  CCGAATTTGA  AAAACTGCTC
    1051  CGCTAA

```

This corresponds to the amino acid sequence <SEQ ID 446; ORF111>:

```

45      1  MPSETRLPNF  IRVLIFALGF  IFLNACSEQT  AQTVTLQGET  MGTITYVKYL
      51  SNNRDKLPSP  AEIQKRIDDA  LKEVNRMQST  YQPDSEISRF  NQHTAGKPLR
     101  ISSDFAHVTA  EAVRLNRLTH  GALDVTVGPL  VNLWGFPGDK  SVTREPSPEQ
     151  IKQAASYTGI  DKIIKQKGD  YASLSKTHPK  AYLDLSSIAK  GFGVDKVAGE
     201  LEKYGIQNYL  VEIGGELHGK  GKNARGEPRW  IGIEQPNIVQ  GGNTQIIVPL
     251  NNRSLATSGD  YRIFHVDKNG  KRLSHIINPN  NKRPISHNLA  SISVVADSAM
     301  TADGLSTGLF  VLGETEALKL  AEREKLAVFL  IVRDKGGYRT  AMSSEFEKLL
     351  R*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) from strain A of *N. meningitidis*:

5	orf111a.pep	10	20	30	40	50	60
	orf111	10	20	30	40	50	60
10	orf111a.pep	70	80	90	100	110	120
	orf111	70	80	90	100	110	120
15	orf111a.pep	130	140	150	160	170	180
	orf111	130	140	150	160	170	180
20	orf111a.pep	190	200	210	220	230	240
	orf111	190	200	210	220	230	240
25	orf111a.pep	250	260	270	280	290	300
	orf111	250	260	270	280	290	300
30	orf111a.pep	310	320	330	340	350	
	orf111	310	320	330	340	350	

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

40	1	ATGCCGTCG	AAACACGCCT	GCCGAACCTT	ATCCGCACCT	TGATATTTGC
	51	CCTGAGTTT	ATCTTCTCTGA	ACGCCTGTTT	GGAACAAACC	GCGCAAACCG
	101	TTACCCGCA	AGGTGAAACG	ATGGGCACGA	CCTATACCGT	CAAATACCTT
	151	TCAAATAATC	GGGACNAACT	CCNCTCACCT	GCCGAAATAC	AAAANCGCAT
45	201	CGATGACGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TATCAGCCCG
	251	ACTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCTCCGC
	301	ATTTCAAGCG	ACTTCGCACA	CGTTACTGCC	GAAGCCGTCC	ACCTGAACCG
	351	CCTGACACAC	GGCGGCTGG	ACGTAACCGT	CGGCCCTTG	GTCAACCTTT
	401	GGGGATTTCG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
50	451	ATCAAACAAG	CAGCATCTTA	TACGGGCATA	GACAAATCA	TTTGTAAACA
	501	AGGCAAAGAT	TACGCTTCTT	TGAGCAAAAC	CCACCCCAAG	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATNANGT	TGCGGGCGAA
	601	CTGGA AAAAT	ACGGCATTCA	AAAT'TATCTG	GTCGAAATCG	GCGGNGAGTT
	651	GCACGGCAA	GNCAAAAACG	CGCGGGCGGA	ACCTTGGCGC	ATCGGCATCG
	701	AACAGCCCAA	CATCGTCCAA	GGCGGCAATA	CGCAGATTAT	CGTCCCGCTG
55	751	AACAACCGTT	CGNTTGCCAC	TTCCGGCGAT	TACCGTATTT	TCCACGTCGA
	801	TAAAAGCGGC	AAACGCCTCT	CCCATATCAT	TAATCCGAAC	AACAAACGAC
	851	CCATCAGCCA	CAACCTCGCC	TCCATCAGCG	TGNTCGCAGA	CAGTGCAGATG
	901	ACGGCGGACG	GCTTNTCCAC	AGGATTATTC	GTATTGGGCG	AAACCGAAGC
	951	CTTAAAGCTG	GCAGAGCGCG	AAAAACTCGC	TGTTTTCTCT	ATTGTACAGG
60	1001	ATAAAGGCGG	CTACCGCACC	GCCATGTCTT	CCGAATTGTA	AAAACGTCTC
	1051	CGCTAA				

This encodes a protein having amino acid sequence <SEQ ID 448>:

1 MPSETRLPNF IRTLIFALSF IFLNACSEQT AQTVTLQGET MGTITYTVKYL  
51 SNNRDXLPSP AEIQXRIDDA LKEVNRQMST YQPDSEISRF NQHTAGKPLR



5

101	ISSDFAHVTA	EAVHLNRLTH	GALDVTVGPL	VNLWGFPGPK	SVTREPSPEQ
151	IKQAASTGT	DKIILKQGGK	YASLSKTHPK	AYLDLSSIAK	GFGVDXVAGE
201	LEQAGIQNYL	VEIGGELHGK	XKNARGEFPWR	IGIEQPNIVQ	GGNTQIIVPL
251	NNRSXATSGD	YRIFHVDKSG	KRLSHIINPN	NKRFISHNLA	SISVXADSAM
301	TADGXSTGLF	VLGETEALKL	AEREKLAVFL	IVRDKGGYRT	AMSSEFEKLL
351	R*				

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N.*

10 *gonorrhoeae*:

			10	20	30	40	50	60
	orf111ng		MPSETRLPNLIRALIFALGFIFLNACSEQTAQTVTTLQGETMGTTYTVKYLNSNNRDKLPSP           :   :   :					
15	orf111		MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTTLQGETMGTTYTVKYLNSNNRDKLPSP 10                20                30                40                50                60					
			70	80	90	100	110	120
	orf111		AKIQKRIDDALKEVNRQMSTYQTDSEISRFNQHTAGKPLRISSDFAHVTAEEAVRLNRLTH   :					
20	orf111		AETIKRIDDALKEVNRQMSTYQPDSEISRFNQHTAGKPLRISSDFAHVTAEEAVRLNRLTH 70                80                90                100                110                120					
			130	140	150	160	170	180
25	orf111ng		GALDVTVGPLVNWLWGFGPDKSVTREPSPEQIKQAASYTGIDKIILQQGKDYPASLSKTHPK           :					
	orf111		GALDVTVGPLVNWLWGFGPDKSVTREPSPEQIKQAASYTGIDKIILKQGKDYPASLSKTHPK 130                140                150                160                170                180					
			190	200	210	220	230	240
30	orf111ng		AYLDLSSIAGFGVDKVAGELEKYGIQNYLVEIGGELHGKGKNHGEHPWRIGIEQPNIHQ           :					
	orf111		AYLDLSSIAGFGVDKVAGELEKYGIQNYLVEIGGELHGKGKNARGEHPWRIGIEQPNIVQ 190                200                210                220                230                240					
35			250	260	270	280	290	300
	orf111ng		GGNTQIIIVPLNNRSLATSGDYRFHFVDKNGKRLSHIINPNNKRPISHNLASISVVSDSAM           :					
	orf111		GGNTQIIIVPLNNRSLATSGDYRFHFVDKNGKRLSHIINPNNKRPISHNLASISVVADSAM 250                260                270                280                290                300					
40			310	320	330	340	350	
	orf111ng		TADGLSTGLFVLGTEALRLAEQEKLAVFLIVRDGDGYRTAMSSEFAKLLRX           :					
45	orf111		TADGLSTGLFVLGTEALKLAEREKLAVFLIVRDGGGYRTAMSSEFEKLLRX 310                320                330                340                350					

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

	1	ATGCCGCTCTG	AAACACGCCT	GCCGAACCTT	ATCCGCGCCT	TGATATTTGC
	51	CCTGGGTTTC	ATCTTCTCTGA	ACGCCTGTTC	GGaacaacC	GCGCAaacCG
50	101	TTACCTCTGCA	AGGCGAAAcg	aTGGGTACGA	CCt'ATACCGT	CAAAATACCTT
	151	TCAAATAACT	GGGACAAACT	CCCTCCCCT	CCGAAATATAC	AAAAGCGCAT
	201	TGATGATGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TACCAGACCG
	251	ATTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCTCCGC
	301	ATTTCAAGCG	ATTCGCACA	CGTTACCGCC	GAAGCCGTCC	GCCTGAACCG
55	351	CCTGACTCAC	GCTGACTGG	ACGTAAACGT	CGGCCCTTTG	CGCTACCTTT
	401	GGGGGTTTCGG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
	451	ATCAAACAGG	CGGCATCTTA	TACGGGCATA	GACAAAATCA	TTTTGCAACA
	501	AGGCAAAAGAT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAAA	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGCG	TTGATAAAGT	TGCGGGCGAA
60	601	CTGGA AAAAT	ACGGCATTCA	AAATTATCTG	GTCGAAAAtcg	gcggcGAGTT
	651	GCACAGCCAAA	GGCAAAAATG	CGCACGGCGA	ACCGTGGCGC	atcgCTATAG
	701	AGCAACCCAA	TATCATCCAA	GgcgGCAaata	CGCAGATTAT	cgtcccgctg
	751	aaCaaccggtt	cgctTGGCAC	TTCCGGCGAT	TAcgctaTTT	tccacgtcgA
	801	TAAAAAcggc	aaacgccttt	cccacaTCAT	CAATCCCaAc	aacAAACgac
65	851	ccATCAGcca	caacctcgcc	tccatcagcg	tggtctcAGA	CAGTGC AATG
	901	ACGGCGGACG	GTTtatCCAC	AGGATTATTT	GT'TTTAGCGC	AAACCGAAGC
	951	CTTAAAGGCTG	CGAGAACAAG	AAAAACTCGC	GT'TTTTCCTA	ATTGTCGGG

-270-

1001 ATAAGGACGG CTACCGCACC GCCATGTCTT CCGAATTGTC CAAGCTGCTC  
1051 CGCTAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

5           1   MPSETRLPNL IRALIFALGF IFLNACSEQT AQTVTLOGET MGTTYTVKYL  
          51   SNNRDKLPSP AKIQKRIDDA LKEVNRQMSY YQTDSEISRF NQHTAGKPLR  
         101   ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFPGDK SVTREPSPEQ  
         151   IKQAASYTGI DKIILOQGGK YASLSKTHPK AYLDLSSIAK GFGVDKVAGE  
         201   LEKYGIQNYL VEIGGELHGK GKNAHGEPWR IGIEQPNIIQ GGNTQIIIVPL  
         251   NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVSDDAM  
10          301   TADGLSTGLF VLGETEALRL AEQEKLAVFL IVRDKDGYRT AMSSEFAKLL  
         351   R\*

This protein shows homology with a hypothetical lipoprotein precursor from *H. influenzae*:

15           sp|P44550|YOJL\_HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR >gi|1074292|pir|4  
          hypothetical protein HI0172 - Haemophilus influenzae (strain Rd KW20)  
          >gi|1573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346  
          Score = 353 bits (896), Expect = 9e-97  
          Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)  
  
20          Query: 7   LPNLIRALIFALGFIFLNACSEQTAQVTTLQGETMGTTYTVKYL SNNRDKLPSPAKIQKR 66  
              +   LI   +I       + L AC ++T + ++L G+TMGTTY VKYL +       S K +  
              Sbjct: 1   MKKLISGIIAVAMALSLAACQKET-KVISLSGKTMGTYHVKYLDGGSITATSE-KTHEE 58  
  
              Query: 67   IDDALKEVNRQMSYQTDSEISRFNQHT-AGKPLRISSDFAHVTA EAVRLNRLTHGALDV 125  
              I+   LK+VN +MSTY+ DSE+SRFNQ+T   P+ IS+DFA V AEA+RLN++T GALDV  
25          Sbjct: 59   IEAILKDVNAKMSYKKDSELSRFNQNTQVNTPIEISADFAKVLAEAIRLNKVTEGALDV 118  
  
              Query: 126   TVGPLVNLWGFPGDKSVTREPSPEQIKQAASYTGIDKIILOQGGKDYASLSKTHPKAYLDL 185  
              TVGP+VNLWGFPG+K   ++P+PEQ+ +   ++ GIDKI L   K+ A+LSK P+ Y+DL  
30          Sbjct: 119   TVGPVNLWGFGEKRPKQPTPEQLAERQAWVGIDKITLDTNKEKATLSKALPQVYVDL 178  
  
              Query: 186   SSIAKGFVGVDKVAGELEKYGIQNYLVEIGGELHGK GKNAHGEPWRIGIEQPNIIQGGNTQ 245  
              SSIAKGFVGVD+VA +LE+   QNY+VEIGGE+   KGKN G+PW+I IE+P       +  
  
              Sbjct: 179   SSIAKGFVGVDQVAEKLEQLNAQNYMVEIGGEIRAKGKNIEGKWPQIAIEKPTTTGERAVE 238  
35          Query: 246   IIVPLNNRSLATSGDYRIHFVDKNGKRLSHIINFNNKRPISHNLASISVVSDDAMTADGL 305  
              ++ LNN +A+SGDYRI+   ++NGKR +H I+P   PI H+LASI+V++ ++MTADGL  
              Sbjct: 239   AVIGLNNMGMASSGDYRIY-FEENGKRFAHEIDPKTGYPIQHHLASITVLAPTSM TADGL 297  
  
              Query: 306   STGLFVLGETEALRLAEQEKLAVFLIVRDKDGYRTAMSSEFAKL 349  
40          STGLFVLGE +AL +AE+   LAV+LI+R +G+ T   SS F KL  
              Sbjct: 298   STGLFVLGEDKALEVAEKNLAVYLIIRTDNGFVTKSSSAFKKL 341

Based on this analysis, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 54

45   The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 451>:

50           1   ..CCGTGCCGCC GACAGGGCGA CGACGTGTAT GCGGCGCACG CGTCCCCTCA  
          51   AAAATTGTGG CTGCGCTTCA TCGGCGGCGG GTCGCATCAA AATATACGGG  
         101   GCGGCGCGGC TCGGCGACGG TGGCGCAAAG GCGTGCAAAT CGGCGCGCAG  
         151   GTGTTGTGAC GGCAAAATGA AGGCAGCCkA yTGGCAATCG GCGTGATGGG  
         201   CGGCAGGGCC GGCCAGCACG CwTCAGTCAA CGGCAAAGGC GGTGCGGCAG  
         251   gCAGTGATT GTATGGTTAT GgCGGGGgTG TTTATGCTgC GTGGCATCAG  
         301   TTGCGCGATA AACAAACGGG TgCGTATTG GACGGCTGGT TGCAATACCA  
         351   ACGTTTCAA CACCGCATCA ATGATGAAAA CCGTGCGGAA CgCTACAAAA  
         401   CCAAAGGTTG GACGGCTTCT GTCGAAGGCG GCTACAACGC GCTTGTGGCG  
55          451   GAAGGCATTG TCGGAAAAGG CAATAATGTG CGGTTTACC TACAACCGCA  
         501   GgCGCAGTTT ACCTACTTGG GCGTAAACGG CGGCTTACC GACAGCGAGG  
         551   GGACGGCGGT CGGACTGCTC GGCAGCGGTC AGTGGCAAAG CCGCGCCGGC  
         601   AtTCGGGCAA AAACCCGTTT TGCTTTGCGT AACGGTGTC ATCTTCAGCC  
         651   TTTTGCCGCT TTTAATGTt TGACAGGTC AAAATCTTTC GCGGTGAAAA  
60          701   TGGACGGCGA AAAACAGACG CTGGCAGGCA GGACGGCACT CGAAGGGCGG

-271-

751 TTCGGTATTG AAGCCGGTTG GAAAGGCCAT ATGTCCGCA..

This corresponds to the amino acid sequence &lt;SEQ ID 452; ORF35&gt;:

1 ..PCRRQGGDDVY AAHASRQKLW LRFIGGRSHQ NIRGGAAADG WRKGVQIGGE  
 51 VFVRQNEGSX LAIGVMGGRA GQHASVNGKG GAAGSDLYGY GGGVYAAWHQ  
 101 LRDQKTGAYL DGWLQYQRFK HRINDENRAE RYKTKGWTAS VEGGYNALVA  
 151 EGIVGKGNNV RFYLQPPAQF TYLGVNGGFT DSEGTAVGLL GSGQWQSRAG  
 201 IRAKTRFALR NGVNLQFFAA FNVLHRSKSF GVEMDGEKQT LAGRTALEGR  
 251 FGIEAGWKGH MSA..

Computer analysis of this amino acid sequence gave the following results:

10 Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

ORF and virg-h protein show 51% aa identity in 261aa overlap:

Orf35 5 QGDDVYAAHASRQKLWLRFIGGRSHQNIIRGGAA-ADGWRKGVQIGGEVFVRQNEGSXLAI 63  
 + D++ R+ LWLR I G S+Q ++G A +G+RKGVQ+GGEVF QNE + L+I  
 15 virg-h 396 KNSDIFDRTLPRKGLWLRVIDGHSNQWVQKKTAPVEGYRKGVQLGGEVFTWQNESNQLSI 455  
 Orf35 64 GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDQKTGAYLDGWLQYQRFKH 121  
 G+MGG+A Q ++ + ++ G+G GVYA WHQL+DKQTGAY D W+QYQRF+H  
 20 virg-h 456 GLMGGQAEQRSTFHNPDTNLTGNVKGFGAGVYATWHQLQDKQTGAYADSWMQYQRFH 515  
 Orf35 122 RINDENRAERYKTKGWTASVEGGYNALVAEGIVGKGNNVRFYLQPPAQFTYLGVNGGFTD 181  
 RIN E+ ER+ +KG TAS+E GYNAL+AE KGN++R YLQPPAQ TYLGVNG F+D  
 25 virg-h 516 RINTEDGTERFTSKGITASIEAGYNALLAEHFTKKGNSLRVYLQPPAQFTYLGVNGKFS 575  
 Orf35 182 SEGTAVGLLGSQWQSRAGIRAKTRFALRNGVNLQFFAAFNVLHRSKSFVEMDGEKQTL 241  
 SE V LLGS Q Q+R G++AK +F+L + ++PFAA N L+ +K FGVEMDGE++ +  
 30 virg-h 576 SENAHVNLLGSRQLQTRGVQAKAQFSLYKNIAIEPFAAVNALYHNKPFVEMDGERRVI 635  
 Orf35 242 AGRTALEGRFGIEAGWKGHMS 262  
 +TA+E + G+ K H++  
 30 virg-h 636 NNKTAIESQLGVAVKIKSHLT 656

Homology with a predicted ORF from *N. meningitidis* (strain A)ORF35 shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) from strain A of *N.**meningitidis*:

35 orf35.pep 10 20 30  
 PCRRQGGDDVYAAHASRQKLWLRFIGGRSHQNIIRG  
 :||||| :||||| :|||||  
 orf35a QRLAIPEAEAVLYAQQAANTLFGRLAADRGDDVYAADPSRQKLWLRFIGGRSHQNIIRG  
 310 320 330 340 350 360  
 40 orf35.pep 40 50 60 70 80 90  
 GAAADGWRKGVQIGGEVFVRQNEGSXLAI GVMGGRAGQHASVNGKGGAAGSDLYGYGGGV  
 :||||| :||||| :||||| :||||| :||||| :|||||  
 45 orf35a GAAADGRRKGVQIGGEVFVRQNEGSRLAI GVMGGRAGQHASVNGKGGAAGSYLHGYGGGV  
 370 380 390 400 410 420  
 orf35.pep 100 110 120 130 140 150  
 YAAWHQLRDQKTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGIV  
 :||||| :||||| :||||| :||||| :||||| :|||||  
 50 orf35a YAAWHQLRDQKTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGVV  
 430 440 450 460 470 480  
 orf35.pep 160 170 180 190 200 210  
 GKGNVRFYLQPPAQFTYLGVNGGFTDSEGTAVGLLGSQWQSRAGIRAKTRFALRNGVN  
 :||||| :||||| :||||| :||||| :||||| :|||||  
 55 orf35a GKGNVRFYLQPPAQFTYLGVNGGFTDSEGTAVGLLGSQWQSRAGIRAKTRFALRNGVN  
 490 500 510 520 530 540  
 orf35.pep 220 230 240 250 260  
 LQPFAAFNVLHRSKSFVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA  
 :||||| :||||| :||||| :||||| :||||| :|||||

orf35a LQPFAAFNVLHRSKSFVEMDGEKQTLAGRRTALEGRFGIEAGWKGHMSARIGYGKRTDGD  
550 560 570 580 590 600

5 orf35a KEAALSLKWLFX  
610 620

The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

```

1  ATGTTTCAGAG CTCAGCTTGG TTCAAATACT CGTTCCTACCA AAATCGGCGA
51  CGATGCCGAT TTTTCATTTT CAGACAAGCC GAAACCCGGC ACTTCCCATT
101 ATTTTTCAG CCGTAAAACC GATCAAAATT CATCCGAATA TGGGTATGAC
151 GAAATCAATA TCCAAGGTAA AAATAACAAT AGCGGCATAC TCGCCGTCGA
201 TAATATGCCC GTTGTTAAGA AATATATTAC AGATACTTAC GGGGATAATT
251 TAAAGGATGC GGTTAAGAAG CAATTACAGG ATTTATACAA AACAAGACCC
301 GAAGCTTGGG AAGAAAATAA AAAACGGACT GAGGAGGCGT ATATAGAACA
351 GCTTGGACCA AAATTTAGTA TACTCAAACA GAAAAACCCC GATTTAATTA
15 401 ATAAATTGGT AGAAGATTCC GTACTCACTC CTCATAGTAA TACATCACAG
451 ACTAGTCTCA ACAACATCTT CAATAAAAAA TTACACGTCA AAATCGAAAA
501 CAAATCCACAC GTCGCCGGAC AGGTGTTGGA ACTGACCAAG ATGACGCTGA
551 AAGATTCCCT TTGGGAACCG CGCCGCCATT CCGACATCCA TATGCTGGAA
601 ACTTCCGATA ATGCCCGCAT CCGCCTGAAC ACGAAAGATG AAAAAGTAC
20 651 CCTCCATAAA GCGTATCAGG GCGGTGCGGA TTTCTGTTC GGCTACGACG
701 TCGCGGAGTC GGCACAAACC GCCCTGACCT TTGAAGAAAA AGTCAGCGGA
751 CAATCCGCGC TGGTTTGGGA ACGCCGGCCG GAAATCTGA AAACGCTCGA
801 CGGCGCAGCA CTGATTGCGG CGGAAAAGGC AGACTCTAAT TCGTTTGCCT
851 TTAAACAAAA TTACCGGCAG GACTGTACG AATTATTGCT CAAGCAATGC
25 901 GAAGGCGGAT TTTGCTTGGG CGTGCAGCGT TTGGCTATCC CCGAGGCGGA
951 ACGGTTTTTA TATGCCAAC AGGCTTATGC GGCAATACT TTGTTCCGGC
1001 TCGGTGCCGC CGACAGGGGC GACGACGTGT ATGCCGCCGA TCGTCCCGT
1051 CAAAAATTGT GGCTGCGCTT CATCGCGGCG CGGTGCGATC AAAATATACG
1101 GGGCGGCGCG GCTGCGGACG GGGCGCGCAA AGGCGTGCAA ATCGGCGGCG
30 1151 AGGTGTTTGT ACGGCAAAAT GAAGGCAGCC GGCTGGCAAT CGGCGTGATG
1201 GCGGGCAGGG CTGGCCAGCA CGCATCAGTC AACGGCAAAG GCGGTGCGGC
1251 AGGCAGTTAT TTGCATGGTT ATGGCGGGGG TGTATTATGCT GCGTGGCATC
1301 AGTTGCGCGA TAAACAAACG GGTGCGTATT TGGACGGCTG GTTGCAATAC
1351 CAACGTTTCA AACACCGCAT CAATGATGAA AACCGTGCGG AACGCTACAA
35 1401 AACCAGAGGT TGGACGGCTT CTGTGGAAGG CGGCTACAAC GCGCTTGTGG
1451 CGGAAGGCGT TGTGCGAAAA GGCAATAATG TGCGGTTTTA CCTGCAACCG
1501 CAGGCGCAGT TTACCTACTT GGGCGTAAAC GGGCGGCTTA CCGACAGCGA
1551 GGGGACGGCG GTCGGACTGC TCGGCAGCGG TCACTGGCAA AGCCGCGCCG
1601 GCGATTCGGC AAAAACCCGT TTTGCTTTGC GTAACGGTGT CAATCTTCAG
40 1651 CCTTTTGCCG CTTTAAATGT TTTGCACAGG TCAAATCTT TCGGCGTGGA
1701 AATGGACGGC GAAAAACAGA CGCTGGCAGG CAGGACGGCG CTCGAAGGGC
1751 GGTTCGGCAT TGAAGCCGGT TGGAAGGCC ATATGTCCGC ACGCATCGGA
1801 TACGGCAAAA GGACGGACGG CGACAAAGAA GCCGCATGT CGTCAAATG
1851 GCTGTTTTGA

```

45 This encodes a protein having amino acid sequence <SEQ ID 454>:

```

1  MFRAQLGSNT RSTKIGDDAD FSFSDKPKPG TSHYFSSGKT DQNSSEYGYD
51  EINIQGNYN SGILAVDNMP VVKYITDTY GDNLKDAVKK QLQDLYKTRP
101  EAWENKRT EEAYIEQLGP KFSILKQKNP DLINKLVES VLTSPHSNTSQ
151  TSLNNIFNKK LHVKIENKSH VAGQVLELTK MTLKDSLWEP RRHSDIHMLE
50 201  TSDNARIRLN TKDEKLTVHK AYQGGADFLF GYDVRESDKP ALTFEEKVSG
251  QSGVVLERRP ENLKTLDGRK LIAAEKADSN SFAFKQNYRQ GLYELLKQC
301  EGGFCLGVQR LAIPEAEAVL YAQQAYAANT LFGLRAADRG DDVYAADPSR
351  QKLWLRFIGG RSHQNIIRGA AADGRRKGVQ IGGEVVRQN EGSRLAIGVM
401  GGRAGQHASY NGKGAAGSY LHGYGGGVYA AWHQLRDKQT GAYLDGWLQY
55 451  QRFKHRINDE NRAERYKTKG WTASVEGGYN ALVAEGVVGK GNNVRFYLPQ
501  QAQFTYLVGN GGFTDSEGTA VGLLGSQWQ SRAGIRAKTR FALRNGVNLQ
551  PFAAFNVLHR SKSFGVEMDG EKQTLAGRRTA LEGRFGIEAG WKGHMSARIG
601  YGKRTDGDKE AALSLKWLFX*

```

Homology with a predicted ORF from *N. gonorrhoeae*

60 ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) from *N. gonorrhoeae*:

```

orf35.pep PCRRQGGDDVYAAHASRQKLWLRFIGGRSHQNIIRG 34
:::|:: |::|::|::|::|::|
orf35ngh FTKVQERDDIAIYAQQAQAANTLFLALRLNDKNSDIFDRTLPRKGLWLRVIDGHSNQWVQG 370

```

5	orf35.pep	GAA-ADGWRKGVQIGGEVFVRQNEGSXLAIGVMGGRAGQHASVNGKG--GAAGSDLYGYG	91
	orf35ngh	KTAPVEGYRKGVQLGGEVFTWQNESNQLSIGLMGGQAEQRSTFRNPDTNLTGNVKGFG	430
10	orf35.pep	GGVYAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAE	151
	orf35ngh	AGVYATWHQLQDKQTGAYVDSWMQYQFRHRINTEYATERFTSKGITASIEAGYNALLAE	490
15	orf35.pep	GIVGKGNVRFYLPQQAQFTYLVNGGGFTDSEGTAVGLLGSQWQSRAGIRAKTRFALRN	211
	orf35ngh	HFTKKGNSLRVYLQPPAQLTYLVNGKFSSENAQVNLLGSRQLQSRVGVQAKAQFAFTN	550
15	orf35.pep	GVNLQPFAAFNVLRHSKSFVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA	263
	orf35ngh	GVTFQPFVAVNSIYQQKPFGEIDGDRRVNNKTVIETQLGVAARIKSHLTLOASFNRQT	610

A partial ORF35ngh nucleotide sequence <SEQ ID 455> is predicted to encode a protein having partial amino acid sequence <SEQ ID 456>:

20	1	..KKLRDRNSEY	WKEETYHIKS	NGRTYPNIPA	LFPKHFPDPP	ENINNSKKIS
	51	FYDKEYTEDY	LVGFARGFGV	EKRNGEEKEP	LRQYFKDCVN	TENSNNNDNCK
25	101	ISSFGNYGPI	LIKSDIFALA	SQIKNSHINS	EILSVGNIE	WLRPTLNKLT
	151	GWQEHLYAGL	DPFHYIEVTD	NSHVIGQITD	LGALELTNSL	WKPRWNSNID
30	201	YLITKNAEIR	FNTKNESLLV	KEDYAGGARF	RFAYDLKDKV	PEIPVLTFEK
	251	NITGTSDIIF	EGKALDNLKH	LDGHQIVKVN	DTADKDAFRL	SSKYRKGITYT
35	301	LSLQQRPEGF	FTKVQERDDI	AIYAQQAQAA	NTLFALRLND	KNSDIFDRTL
	351	PRKGLWLRVI	DGHSNQWVQG	KTAPVEGYRK	GVQLGGEVFT	WQNESNQLSI
40	401	GLMGGQAEQR	STFRNPDTDN	LTTGNVKGFG	AGVYATWHQL	QDKQTGAYVD
	451	SWMQYQFRFH	RINTEYATER	FTSKGITASI	EAGYNALLAE	HFTKKGNSLR
45	501	VYLQPPAQLT	YLVNGKFS	SENAQVNLLG	SRQLQSRVGV	QAKAQFAFTN
	551	GVTFQPFVAV	NSIYQQKPF	GEIDGDRVI	NNKTVIETQL	GVAARIKSHL
	601	TLQASFNRQT	SKHHHAKQGA	LNLQWTF*		

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 55

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 457>:

40	1	..GCCGAATATG	TTCAGTTCTC	TATAGATTG	TTCAGTGTGG	GTAAATCGGG
	51	GGGCGGTATA	CCTAAGGCTA	AGCCTGTGTT	TGATGCGAAA	CCGAGATGGG
45	101	AGGTTGATAG	GAAGCTTAAT	AAATTGACAA	CTCGTGAGCA	GGTGGAGAAA
	151	AATGTTGAGG	AAACGAGAG	AAGGAGTCAG	AGTAGTCAGT	TTAAAGCCCA
50	201	TGCGCAACGA	GAATGGGAAA	ATAAAACAGG	GTTAGATTTT	AATCATTTTA
	251	TAGGTGGTGA	TATCAATAAA	AAAGGCACAG	TAACAGGAGG	GCATAGCTCA
55	301	ACCCGTGGTG	ATGTACGGGT	GATACAACAA	ACCTCGGCAC	CTGATAAACA
	351	TGGGGT.TTA	TCAAGCGACA	GTGGAAATTN	A	

This corresponds to the amino acid sequence <SEQ ID 458; ORF46>:

45	1	..AEYVQFSIDL	FSVGKSGGGI	PKAKPVFDAK	PRWEVDKRLN	KLTTREQVEK
	51	NVQETRRRSQ	SSQFKAHAQR	EWENKTGLDF	NHFIGGDINK	KGTVTGGHSL
	101	TRGDVRVIQQ	TSAPDKHGXL	SSDSGNX		

Further work revealed further partial nucleotide sequence <SEQ ID 459>:

50	1	..GCAGTGTGCC	TnCCGATGCA	TGCACACGCC	TCAaATTG	CAAACGATTC
	51	TTTTATCCGG	CAGGTTCTCG	ACCGTCAGCA	TTTCGAACCC	GACGGGAAAT
55	101	ACCACCTATT	CGGCAGCAGG	GGGGAACCTG	CCGAGCGCCA	GTCTCATATC
	151	GGATTGGGAA	AAATACAAAG	CCATCAGTTG	GGCAACCTGA	TGATTCAACA
60	201	GGCGGCCATT	AAAGGAAATA	TGGGCTACAT	TGTCCGCTTT	TCCGATCAGC
	251	GGCACGAAGT	CCATTCCCCs	TTTCGACAACC	ATGCCTCACA	TTCGATCTCT
65	301	GATGAAGCCG	GTAGTCCCGT	TGACGGATTT	AGCCTTTACC	GCATCCATTG
	351	GGACGGATAC	GAACACCATC	CCGCCGACGG	CTATGACGGG	CCACAGGGCG
	401	GCGGCTATCC	CGCTCCCAAA	GGCGCGAGGG	ATATATACAG	TTACGACATA

-274-

5 451 AAAGGCGTTG CCCAAATAT CCGCCTCAAC CTGACCGACA ACCGCAGCAC  
501 CGGACAACGG CTTGCCGACC GTTCCACAA TGCCGGTAGT ATGCTGACGC  
551 AAGGAGTAGG CGACGGATTC AAACGCGCCA CCCGATACAG CCCCAGCTG  
601 GACAGATCGG GCAATGCCGC CGAAGCCTTC AACGGCACTG CAGATATCGT  
651 TAAAAACATC ATCGGCGCTG CAGGAGAAAT TGT

This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

10 1 ..AVCLPMHAHA SXLANDSFIR QVLDROHFEP DGKYHLFGSR GELAERQSHI  
51 GLGKIQSHQL GNLMIQAAI KGNIGYIVRF SDHGHEVHSP FDNHASHSDS  
101 DEAGSPVDGF SLYRIHWDGY EHHPADGYDG PQGGGYPAFK GARDIYSYDI  
151 KGVAQNIRLN LTDNRSTGOR LADREFHNAGS MLTQGVGDGF KRATRYSPDL  
201 DRSGNAAEAF NGTADIVKNI IGAAGEI

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N.*

15 *gonorrhoeae*:

orf46.pep AEYVQFSIDLFSVGKSGGGIPKAKPVFDAKPRWEVDRLNKLTTTR 45  
orf46ng PKTGVPFDGKGFPNFEKHVKYDTKLDIQELSGGGIPKAKPVFDAKPRWEVDRLNKLTTTR 217  
20 orf46.pep EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDV 105  
orf46ng EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGAVTGGHSLTRGDV 277  
orf46.pep RVIQQTSA PDKHGXLSSDSGN 126  
25 orf46ng RVIQQTSA PDKHGVLSSDSGN 298

A partial ORF46ng nucleotide sequence <SEQ ID 461> is predicted to encode a protein having partial amino acid sequence <SEQ ID 462>:

30 1 ..RRLKHCHAR LGSAFHRKQD GAHQRFGRYG ATQRLCRSSH PRLGSPKPOC  
51 RTRHRSRQQY LYGSHPHQRD WSCPGKIQLG RHHGTSCRAV ADXRDRICER  
101 EIRRQRQXCR CRLGKIPSL IPKYPLKLEQ RYKKNITSS TVPPSNGKNV  
151 KLADQRHPKT GVPFDGKGFP NFEKHVKYDT KLDIQELSGG GIPKAKPVFD  
201 AKPRWEVDRK LNKLTTRQV EKNVQETRRR SQSSQFKAHA QREWENKTGL  
251 DFNHFIGGDI NKKGAVTGGH SLTRGDV RVI QQTSA PDKHG VLSSDSGN\*

35 Further work revealed the complete gonococcal DNA sequence <SEQ ID 463>:

1 TTGGGCATT CCCGCAAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG  
51 CCTGCCGATG CATGCACACG CCTCAGATTG GGCAACGAT CCCTTTATCC  
101 GgCaggttcT CGaccGTcAG CATTTCGaac ccgacggGaa ATACCaCCTA  
151 TtcggCaGCA GGGGGGAGCT TgcnagcGC aacggccATa tcggattggG  
40 201 aaacaTAcAa Agccatcagt tGggccacct gatgattcaa caggcggccg  
251 ttgaaggaaA TAtcgGctac attgtccgct tttccgatca cgggcacaaa  
301 ttccattcgc ccttcGAcAa ccaTGCCTCA CATTCCGATT CTGACGAAGC  
351 CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CCGCATCCAT TGGGACGGAT  
401 ACGAACACCA TCCCGCCGAC GGCTATGACG GGCCACAGGG CGGCGGCTAT  
45 451 CCGCTCCCA AAGGCGCGAG GGATATATAC AGCTACGACA TAAAAGGCGT  
501 TGCCCAAAAT ATCCGCTCA ACCTGACCGA CAACCGCAGC ACCGGACAAC  
551 GGCTTGCCGA CCGTTTCCAC AATGCCGGCG CTATGCTGAC GCAAGGAGTA  
601 GGGACGGAT TCAAACGCGC CACCCGATAC AGCCCGAGC TGGACAGATC  
651 GGGCAATGcC gccGAAGCCT TCAACGGCAC TGCAGATATC GTCAAAAACA  
50 701 TCATCGGCGC GGCAGGAGAA ATTGTCGGCG CAGGCGATGC CGTGCagGGT  
751 ATAAGCGAAG GCTCAAACAT TGCTGTCATG CACGGCTTGG GTCTGCTTTC  
801 CACCGAAAAC AAGATGGCGC GCATCAACGA TTGGCAGAT ATGGCGCAAC  
851 TCAAAGACTA TGCCGAGCA GCCATCCGCG ATTGGGCAGT CCAAAACCCC  
901 AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA TGGCAGCCAT  
55 951 CCCCATCAAA GGGATTGGAG CTGTCCGGGG AAAATACGGC TTGGGCGGCA  
1001 TCACGGCACA TCCTGTCAAG CGGTGCGAGA TGGGCGCGAT CGCATTGCCG  
1051 AAAGGGAAAT CCGCCGTCAG CGACAATTTT GCCGATGCGG CATACGCCAA  
1101 ATACCCGTCC CCTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC

-275-

5  
10  
15

```

1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGC
1201 AAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT
1251 TGACGGTAAA GGGTTTCCGA ATTTTGAGAA GCACGTGAAA TATGATACGA
1301 AGCTCGATAT TCAAGAATTA TCGGGGGGCG GTATACCTAA GGCTAAGCCT
1351 GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT
1401 GACAACCTCGT GAGCAGGTGG AGAAAAATGT TCAGGAAACG AGAAGAAGGA
1451 GTCAGAGTAG TCAGTTTAAA GCCCATGCGC AACGAGAATG GGAAAAATAA
1501 ACAGGGTTAG ATTTTAATCA TTTTATAGGT GGTGATATCA ATAAGAAAGG
1551 CACAGTAACA GGAGGGCATA GTCTAACCCG TGGTGATGTA CGGGTGATAC
1601 AACAAACCTC GGCACCTGAT AAACATGGGG TTTATCAAGC GACAGTGGAA
1651 ATTA AAAAGC CTGATGGAAG TTGGGAGGTG AAAACGAAAA AAGGTGGGAA
1701 AGTGATGACC AAGCACACCA TGTTCCTAAA AGATTGGGAT GAGGCTAGAA
1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT
1801 AATAAATGGC AGGGTACAAG TAAATCGGGT ATTAAATAG AAGGATTTC
1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

```

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

20  
25  
30

```

1  LGISRKISLI  LSILAVCLPM  HAHASDLAND  PFIRQVLDRO  HFEPDGKYHL
51  FGSRGELAXR  NGHIGLGNIQ  SHQLGHLMIQ  QAAVEGNIGY  IVRFSDHGHK
101 FHSFPDNHAS  HSDSDEAGSP  VDGFSLYRIH  WDGYEHPAD  GYDGPQGGGY
151 PAPKGARDIY  SYDIKGVAQN  IRLNLTDNRS  TQRLADRFH  NAGAMLTQGV
201 GDGFKRATRY  SPELDRSGNA  AEAFTGTADI  VKNIIGAAGE  IVGAGDAVQG
251 ISEGSNIAVM  HGLGLLSTEN  KMARINDLAD  MAQLKDYAAA  AIRDWAVQNP
301 NAAQGIEAVS  NIFMAAIPIK  GIGAVRGKYG  LGGITAHVPK  RSQMGAIALP
351 KGKSAVSDNF  ADAAYAKYPS  PYHSRNIRSN  LEQRYGKENI  TSSTVPPSNG
401 KNVKLADQRH  PKTGVPFDGK  GFNFKEKHVK  YDTKLDIQEL  SGGGIPKAKP
451 VFDAKPRWEV  DRKLNKLTTT  EQVEKNVQET  RRRSQSSQFK  AHQREWENK
501 TGLDFNHFIG  GDINKKGTVT  GGHSLTRGDV  RVIQQTSPD  KHGVYQATVE
551 IKKPDGSWEV  KTKKGKQVMT  KHTMFPKDW  EARIRAEVTS  AWESRIMLKD
601 NKWQGTSGSG  IKIEGFTEPN  RTAYPIYE*

```

30 ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

35  
40  
45  
50  
55  
60

```

orf46-1.pep          10      20      30      40
                    AVCLPMHAHASXLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER
orf46ng-1            LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR
                    10      20      30      40      50      60

orf46-1.pep          50      60      70      80      90      100
                    QSHIGLGKIQSHQLGNLMIQQAIAIKGNIGYIVRFSDHGHEVHSPFDNHASHSDSDEAGSP
orf46ng-1            NGHIGLGNIQSHQLGHLMIQQAIAVEGNIGYIVRFSDHGHKFHSPFDNHASHSDSDEAGSP
                    70      80      90      100      110      120

orf46-1.pep          110     120     130     140     150     160
                    VDGFSLYRIHWDGYEHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRLNLTDNRS
orf46ng-1            VDGFSLYRIHWDGYEHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRLNLTDNRS
                    130     140     150     160     170     180

orf46-1.pep          170     180     190     200     210     220
                    TQRLADRFHNAGSMLTQGVGDGFKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE
orf46ng-1            TQRLADRFHNAGAMLTQGVGDGFKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE
                    190     200     210     220     230     240

orf46-1.pep          I
orf46ng-1            IVGAGDAVQGISSEGSNIAVMHGLGLLSTENKMARINDLADMAQLKDYAAAAAIRDWAVQNP
                    250     260     270     280     290     300

```

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of *N. meningitidis*:

10 20 30 40 50 60

-276-

	orf46a.pep	LGISRKISLILSLAVCLPMHAHASDLANDSFIRQVLDQRHFEPDGKYHLFGSRGELAER
	orf46ng-1	LGISRKISLILSLAVCLPMHAHASDLDNPFIRQVLDQRHFEPDGKYHLFGSRGELAXR
5		10 20 30 40 50 60
	orf46a.pep	SGHIGLGNIQSHQLGNLFIQQAATKGNIGYIVRFSHDHGHEVHSPFDNHASHSDSDEAGSP
10	orf46ng-1	NGHIGLGNIQSHQLGHLMIQQAAVEGNIGYIVRFSHDHGKHFHSPFDNHASHSDSDEAGSP
		70 80 90 100 110 120
15	orf46a.pep	VDGFSLYRIHWDDGYEHHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRLNLTNRS
	orf46ng-1	VDGFSLYRIHWDDGYEHHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRLNLTNRS
		130 140 150 160 170 180
20	orf46a.pep	TGQRLVDRFHNTGSMLTQGVGDGFKRATRYSPELDRSGNAAEFNGTADIVKNIIGAAGE
	orf46ng-1	TGQRLADRFHNTGAMLTQGVGDGFKRATRYSPELDRSGNAAEFNGTADIVKNIIGAAGE
		190 200 210 220 230 240
25	orf46a.pep	IVGAGDAVQGISSEGSNIAVMHGLGILLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
	orf46ng-1	IVGAGDAVQGISSEGSNIAVMHGLGILLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
		250 260 270 280 290 300
30	orf46a.pep	NAAQGIEAVSNIFTAVIPVKIGIAVRGKYGLGGITAHFVKRSQMGAIALPKGKSAVSDNF
	orf46ng-1	NAAQGIEAVSNIFMAAIPKIGIAVRGKYGLGGITAHFVKRSQMGAIALPKGKSAVSDNF
		310 320 330 340 350 360
35	orf46a.pep	ADAAYAKYPSPYHSRNRNIRSNLEQRYGKENITSSTVPPSNGKNVKLANKRHPKTKVPFDGK
40	orf46ng-1	ADAAYAKYPSPYHSRNRNIRSNLEQRYGKENITSSTVPPSNGKNVKLADQRHPKTKVPFDGK
		370 380 390 400 410 420
45	orf46a.pep	GFPNFEKDVKYDTRINTAVPQVN----PIDEPVFN--PKGSVGSASHWSITARIQYAKLP
	orf46ng-1	GFPNFEKHVKYDTKLD--IQELSGGGIPKAKPVFDAKPERWEVDKRLN-KLTTRREQVEKNV
		430 440 450 460 470
50	orf46a.pep	RQGRIRYIPPKNYSAPSAPLPKGPNNGYLDKFGNEWTKGPSRTKGQEFEDVQLSKTGREQ
	orf46ng-1	QETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDVVRVIQOTS
		480 490 500 510 520 530

The complete length ORF46a DNA sequence <SEQ ID 465> is:

55	1	TTGGGCATTT	CCCGCAAAAT	ATCCCTTATT	CTGTCCATAC	TGGCAGTGTG
	51	CCTGCCGATG	CATGCACACG	CCTCAGATTT	GGCAAACGAT	TCTTTTATCC
	101	GGCAGGTTCT	CGACCGTCAG	CATTTCGAAC	CCGACGGGAA	ATACCACCTA
	151	TTCCGCAGCA	GGGGGGAAGT	TGCCGAGCGC	AGCGGTGATA	TCGGATTGGG
	201	AAACATACAA	AGCCATCAGT	TGGGCAACCT	GTTTCATCCAG	CAGGCGGCCA
60	251	TTAAAGGAAA	TATCGGCTAC	ATTGTCCGCT	TTCCGATCA	CGGGCAGCAA
	301	GTCCATTCCC	CCTTCGACAA	CCATGCCTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTCCC	GTTGACGGAT	TCAGCCTTTA	CCGCATCCAT	TGGGACGGAT
	401	ACGAACACCA	TCCCGCCGAC	GGCTATGACG	GGCCACAGGG	CGGCGGCTAT
	451	CCCGCTCCCA	AAGGCGCGAG	GGATATATAC	AGCTACGACA	TAAAAGCGCT
65	501	TGCCCAAAAT	ATCCGCCTCA	ACCTGACCGA	CAACCGCAGC	ACCGGACAAC
	551	GGCTTGTCGA	CCGTTTCCAC	AATACCGGTA	GTATGCTGAC	GCAAGGAGTA
	601	GGCGACGGAT	TCAAACGCGC	CACCCGATAC	AGCCCCGAGC	TGGACAGATC
	651	GGGCAATGCC	GCCGAAGCTT	TCAACGGCAC	TGCAGATATC	GTCAAAAACA
	701	TCATCGGCGC	GGCAGGAGAA	ATTGTCCGCG	CAGGCGATGC	CGTGCAGGGT
	751	ATAAGCGAAG	GCTCAAACAT	TGCTGTTATG	CACGGCTTGG	GTCTGCTTTC
70	801	CACCGAAAAC	AAGATGGCGC	GCATCAACGA	TTTGGCAGAT	ATGGCGCAAC



5  
 10  
 15

```

851 TCAAAGACTA TGCCGCAGCA GCCATCCGCG ATTGGGCAGT CCAAAACCCC
901 AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA CGGCAGTCAT
951 CCCCCTCAA GGGATTGGAG CTGTTGCGGG AAAATACGGC TTGGGCGGCA
1001 TCACGGCACA TCCTGTCAAG CGGTCGCAGA TGGGCGAGAT CGCATTGCCG
1051 AAAGGGAAAT CCGCCGTCAG CGACAAATTTT GCCGATGCGG CATACGCCAA
1101 ATACCCGTC CTTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC
1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGA
1201 AAGAATGTGA AACTGGCAAA CAAACGCCAC CCGAAGACCA AAGTGCCGTT
1251 TGACGGTAAA GGGTTTCCGA ATTTTGAAAA AGACGTAAAA TACGATACGA
1301 GAATTAATAC CGCTGTACCA CAAGTGAATC CTATAGATGA ACCCGTCTTT
1351 AATCCTAAAG GTTCTGTCGG ATCGGCTCAT TCTTGGTCTA TAACTGCCAG
1401 AATTCAATAC GCAAAATTAC CAAGGCAAGG TAGAATCAGA TATATCCAC
1451 CTAAAATTA CTCTCCTTCA GCACCGCTAC CAAAAGGACC TAATAATGGA
1501 TATTTGGATA AATTGGTAA TGAATGGAAT AAAGGTCCAT CAAGAATAA
1551 AGGTCAAGAA TTTGAATGGG ATGTTCAATT GTCTAAAACA GGAAGAGAGC
1601 AACTTGGATG GGCTAGTAGG GATGGTAAGC ATTTAAATAT ATCAATTGAT
1651 GGAAAGATTA CACACAAATG A
  
```

This corresponds to the amino acid sequence <SEQ ID 466>:

20  
 25  
 30

```

1 LGISRKISLI LSILAVCLPM HAHASDLAND SFIRQVLDRO HFEPDGKYHL
51 FGSRGELAER SGHIGLGNIQ SHQLGNLFIQ QAAIKGNIGY IVRFSDHGHE
101 VHSFPDNHAS HSDESDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY
151 PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLVDRFH NTGSMILTQGV
201 GDGFKRATRY SPELDRSGNA AEAFTNGTADI VKNIIGAAGE IVGAGDAVQO
251 ISEGSNIIVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
301 NAAQGIEAVS NIFTAVIPVK GIGAVRGKYG LGGITAHVPK RSQMGEIALP
351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
401 KNVKLANKRH PKTKVPFDGK GFNFEEKDVK YDTRINTAVP QVNPIDEVPF
451 NPKGSVGSAA SWSITARIQY AKLPRQGRIR YIPPKNYSPS APLPKGPNNG
501 YLDKFGNEWT KGPSRTKGQE FEWDVQLSKT GREQLGWASR DGKHLNISID
551 GKITHK*
  
```

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 35 Example 56

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

40

```

1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCTT
51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GCGGTTTCGT
201 CAAAATTGCC GCGGTATTGG CGTTTTGGCT GCGGTTTTTG TTTGACGGGC
251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
351 CGGGCTG...
  
```

45 This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

```

1 MNIHTLLSKQ WTLPPFLPKR LLSLLILLA PNAVFWVLAL LTATARPIVN
51 LDYLPALLI ALPWRFVKIA GVLAFLAVL FDGLMMVIQL PPFMDLIGAI
101 NLVPFILTAP APYQIMTGL...
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

50  
55

```

1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCTT
51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GCGGTTTCGT
201 CAAAATTGCC GCGGTATTGG CGTTTTGGCT GCGGTTTTTG TTTGACGGGC
251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
351 CGGGCTGTTG CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
  
```

-278-

5  
10  
15  
20

```

401 CCGCCGCCAA AACCGACTTC CGGCACATTG CCGTCTGCGC CGCCGTTGTG
451 GCGGCAGCCG GCTATTTTAC CGGCCATTG AGTTACTACG ACCGGGGTCCG
501 GATGGCCAAT ATCTTCGGCG CAAACAACCTT CTACTACGCC AAAAGTCAGG
551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
601 GTCGATCCCG TCTTCTCTCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
701 GGGGGCTGCC GGCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
801 CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCCG
901 TGCCTCCCA ACCGTTTGAA ACAAGAAGT TACGCCACCT TTGCGATGCA
951 CCGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001 GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC
1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTCCGGC AAGTGTCCGG
1101 ATTTTTCAAA AAACACGACA AGGGACTGTT TTAAGTATG ACGCTGACCA
1151 GCCAGCCGA CTATCCCGAA TCCGACATTT TCAACCACAG GCTCAAATGC
1201 ACCGAATATG GCCTGCCCGC CGAAACCGAC CTCTGCCGCA ATTTCAGCCT
1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCCGAAA
1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCGCC GCCCGTCGGC
1351 AACCTCAATG AAACCTTCCG CTACCTCAAA CAGGGGCACG TCGCCTGGCT
1401 GAACTTCAAA ATCAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 470; ORF48-1>:

25  
30

```

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFWLAL LTATARPIVN
51 LDYLPALLI ALPWRVKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIAVCAAVV
151 AAGYFTGHL SYDRGRMAN IFGANFYA KSQAMLYTVS QNADFITAGL
201 VDPVFLPLGN QQRAATHLNE PKSQKILFIV AESWGLPANP ELQNAFALK
251 LAQKDRFSVW ESGSFPIGA TVEGEMRELC AYGLRGFAL RRAPEKFKAR
301 CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQEIKT AENLIGKKTC
351 AIFGGVCDSE LFGEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHLKC
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPVPG
451 NLNETFRYLK QGHVAWLNFK IK*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF48 shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) from strain A of *N. meningitidis*:

40  
45  
50

```

          10      20      30      40      50      60
orf48.pep MNIHTLLSKQWTLPPFLPKRLLLSLILLAPNAVFWLALLTATARPIVNLDYLPALLI
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf48a    MNIHTLLSKQWTLPPFLPKRLLLSLILLXPNVFWLALLTATARPIVNLXYLPALLI
          10      20      30      40      50      60

          70      80      90      100     110     119
orf48.pep ALPWRVKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf48a    ALPWRXVKIXGLAXWLAVLFDGLMMVIQLFPFMDLIGAINLVPFIXTAPALYQIMTGLL
          70      80      90      100     110     120

orf48a    LLYMLAMPFVLQKAAKTDFRHIAACA AVVVAAGYFTGHL SXDRGRMANIFGANFYA
          130     140     150     160     170     180

```

The complete length ORF48a nucleotide sequence <SEQ ID 471> is:

55  
60

```

1 ATGAATATTC ACACCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCTT
51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTNNCC CCCAATGCGG
101 TGTGTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TTGGANTACC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCGTNTCGT
201 CAAAATTGNC GGCGTATTGG CGTNTGGCT GGCGGTTTGG TTTGACGGGC
251 TGATGATGGT GATCCAACCT TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CCTTCATCNT GACCGCCCC GCCCTTTATC AGATAATGAC
351 CGGGCTGTTA CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
401 CCGCCGCCAA AACCGACTTC CGACACATTG CCGCCTGTGC CGCGTTGTG
451 GTGGCAGCCG GCTATTTTAC CGGCCATTG AGTTANTACG ACCGGGGGCG

```

-279-

501 GATGGCCAAT ATCTTCGGCG CAAACAACCTT CTATTACGCC AAAAGTCAGG  
 551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG  
 601 GTCGATCCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA  
 651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT  
 701 GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG  
 751 CTGGCGCAAA AAGANCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT  
 801 CATCGGCGCG ACAGTCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG  
 851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCGC  
 901 TGCTCCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA  
 951 CGGCGCGGGC AGTTGCGCTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG  
 1001 GCTTTCAAGA AATCAAAACC GCCGAAACC TGATCGGTAA AAAAACCTGC  
 1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTCCGCG AAGTGTGCGC  
 1101 ANTTTTCAA AAACACGACA AGGGACTGTT TTAAGGATG ACGCTGACCA  
 1151 GCCACGCCGA CTATCCCGAA TCNGACATT TCAACCACAG GCTCAAATGC  
 1201 ACCGAATATG GCCTGCCCGC CGAAACCGAC NTCTGCCGCA ATTTTCAGCT  
 1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCCGAAA  
 1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC  
 1351 AACCTCAATG AAACCTTCCG CTACCTCAAA CAGGGGCACG TCGNCTGGCT  
 1401 GAACTTCAAA ATCAAATAA

20 This encodes a protein having amino acid sequence <SEQ ID 472>:

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLX PNAVFWVLAL LTATARPIVN  
 51 LXYLPAALLI ALPWRXVKIX GVLAXWLAVL FDGLMMVIQL FPFMDLIGAI  
 101 NLVPFXTAP ALYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIAACAIVV  
 151 VAAGYFTGHL SXYDRGRMAN IFGANFYFA KSQAMLYTVS QNADFITAGL  
 201 VDPVFLPLGN QQRAATHLNE PKSQKILFIV AESWGLPANP ELQNATFAKL  
 251 LAQKXRFVW ESGSFPIGAT IEIEMRELC AYGGRLGFAL RRAPDEKFA  
 301 CLPNRLKQEG YATFAMHGAG SLYDRFSWY PRAGFQEIKT AENLIGKKT  
 351 AIFGGVCDSE LFGVSAFXK KHDKGLFYWM TLTSHADYPE SDIFNHLKLC  
 401 TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPVPV  
 451 NLNETFRYLK QGHVXWLNFK IK\*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

		10	20	30	40	50	60
35	orf48a.pep	MNIHTLLSKQWTLPPFLPKRLLLSLILLXPN	AVFWVLALLTATARPIVN	LXYLPAALLI			
	orf48-1	MNIHTLLSKQWTLPPFLPKRLLLSLILLAPNA	FWVLALLTATARPIVNDY	LPAALLI			
		10	20	30	40	50	60
40	orf48a.pep	ALPWRXVKIXGVLAXWLAVLFDGLMMVIQL	FPFMDLIGAINLV	PFIXTAPALYQIMTG			
	orf48-1	ALPWRXVKIXGVLAXWLAVLFDGLMMVIQL	FPFMDLIGAINLV	PFIXTAPALYQIMTG			
		70	80	90	100	110	120
45	orf48a.pep	LLYMLAMPFVLQKAAKTDFRHIAACA	AVVVAAGYFTGHL	SXYDRGRMANIF	GANFYFA		
	orf48-1	LLYMLAMPFVLQKAAKTDFRHIAVCA	AVVVAAGYFTGHL	SYDRGRMANIF	GANFYFA		
		130	140	150	160	170	180
50	orf48a.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGN	QQRAATHLNEPKS	QKILFIVAESWGL	PANP		
	orf48-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGN	QQRAATHLNEPKS	QKILFIVAESWGL	PANP		
		190	200	210	220	230	240
55	orf48a.pep	ELQNATFAKLLAQKXRFVWESGSFPFI	GATIEGEMRELCAY	GGRLGFALRRAP	DEKFA		
	orf48-1	ELQNATFAKLLAQKXRFVWESGSFPFI	GATIEGEMRELCAY	GGRLGFALRRAP	DEKFA		
		250	260	270	280	290	300
60	orf48a.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFS	WYPRAGFQEIKT	AENLIGKKTCAI	FGGVCDSE		
	orf48-1	CLPNRLKQEGYATFAMHGAGSSLYDRFS	WYPRAGFQEIKT	AENLIGKKTCAI	FGGVCDSE		
		310	320	330	340	350	360
65	orf48a.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFS	WYPRAGFQEIKT	AENLIGKKTCAI	FGGVCDSE		
	orf48-1	CLPNRLKQEGYATFAMHGAGSSLYDRFS	WYPRAGFQEIKT	AENLIGKKTCAI	FGGVCDSE		

-280-

		370	380	390	400	410	420
	orf48a.pep	LFGEVSAXFKKHKDGLFYWMTLTSHADYPESDIFNHRLKCTEYGLPAETDXCRNFSLHTQ					
5	orf48-1	LFGEVSAFFKKHKDGLFYWMTLTSHADYPESDIFNHRLKCTEYGLPAETDLCRNFSLHTQ					
		370	380	390	400	410	420
	orf48a.pep	FFDQLADLIQRPEMKGTEVIIVGDHPPVGNLNETFRYLKQGHVXWLNFKIKX					
10	orf48-1	FFDQLADLIQRPEMKGTEVIIVGDHPPVGNLNETFRYLKQGHVAWLNFKIKX					
		430	440	450	460	470	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.*

15 *gonorrhoeae*:

	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPVNL DYLP AALLI	60
	orf48ng	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPVNL DYLP AALLI	60
20	orf48.pep	ALPWRVFKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL	119
	orf48ng	ALPWRVFKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

25	1	MNIHALLSEQ	WTLPPFLPKR	LLLSLLILLA	PNAVFWVLAL	LTATARPVNL
	51	LDYLP AALLI	ALPWRVFKIA	GVLAFLAVL	FDGLMMVIQL	FPFMDLIGAI
	101	NLVPFILTAP	APYQIMTGLL	LLYMLAMPFV	LQKAAVKTDF	RHIAVCAAVV
	151	AAARYFTGPF	ELLRTGGRWQ	YVQHRRLLS	GSRAFRRRQ	KADVLRRRLGN
	201	PYASMGNGG	.	.	.	.

30 Further work identified the complete gonococcal DNA sequence <SEQ ID 475>:

	1	ATGAATATTC	ACGCCCTGCT	CTCCGAACAA	TGGACGCTGC	CGCCATTCTCT
	51	GCCGAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTGGCC	CCCAATGCGG
	101	TGTTTTGGGT	TTTGCACTG	CTGACCGCCA	CCGCCGCGCC	GATTGTCAAT
35	151	TTGGACTACC	TTCCCGCCGC	GCTGCTGATC	GCCCTGCCTT	GGCGTTTCGT
	201	CAAAATTGCC	GGCGTATTGG	CGTTTTGGCC	GGCGGTTTTG	TTTGACGGGC
	251	TGATGATGGT	GATCCAACCT	TCCCTTTTA	TGGACCTCAT	CGCGGCCATC
	301	AACCTCGTCC	CCTTCATCCT	GACCGCCCC	GCCCCTTATC	AGATAATGAC
	351	CGGCTGTGTG	CTGCTGTATA	TGCTGGCGAT	GCCGTTTGTG	TTGCAAAAAG
40	401	CCGCCGTCAA	AACCGACTTC	CGACACATTC	CCGCTGTGTG	CGCCGTTGTG
	451	CGGCGAGCCG	GCTATTTTAC	CGGCCATTTC	AGTTACTACG	ACCGGGGGCG
	501	GATGGCCAAT	ATCTTCGGCG	CAACAACCTT	CTATTACGCG	aAAAGTCAGG
	551	CGATGCTCTA	CACCGTCAGC	CAGAATGCCG	ACTTTATTAC	CGCCGgcctG
	601	GTCGACCCCG	TCTTCCTCCC	CTTGGGCAAT	CAGCAGCGTG	CCGCCACGCG
	651	GCTGAGTGAG	CCGAAATCTC	AAAAATCCT	CTTTATCGTC	GCCGAATCTT
45	701	GGGGGCTGCC	GGGCAATCCC	GAGCTTCAAA	ACGCCACTTT	TGCCAAACTG
	751	CTGGCGCAAA	AAGACCGTTT	TTCGGTTTGG	GAAAGCGGCA	GTTTTCCCTT
	801	CATCGGCGCG	ACGGTCGAAG	GCGAAATGCG	CGAATTGTGC	GCCTACGGCG
	851	GTTTGCGCGG	GTTTCGACTG	CGCCGCGCGC	CCGACGAAAA	ATTTGCCCGC
	901	TGCCTCCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TTGCGATGCA
50	951	CGGCGCGGGT	AGTTCGCTTT	ACGACCGCTT	CAGCTGGTAT	CCGAGGGCGG
	1001	GCTTTCAAAA	AATCAAAACC	GCCGAAAACC	TGATCGGTAA	AAAAACCTGC
	1051	GCCATTTTCG	GCGGCGTGTG	CGACAGCGAG	CTGTTTCGGCG	AAGTGTGCGG
	1101	ATTTTTCAAA	AAACACGACA	AGGGACTGTT	TTACTGGATG	ACGCTGACCA
	1151	GCCACGCCGA	CTATCCCGAA	TCCGACATTT	TCAACCACAG	GCTCAAATGC
55	1201	ACCGAATACG	GCCTGCCCCG	CGAAACCGAC	CTCTGCCGCA	ATTTCAGCCT
	1251	GCACACCCAA	TtcttcgACC	AACTGGCGGA	TTTGATCCGA	CGCCCCGAAA
	1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCGCGC	ACCATCCGCG	GCCCCGTCGC
	1351	AACCTCAATG	AAACCTCCG	CTACCTCAAA	CAGGGACACG	TCGCTTGCTT
	1401	GCACTTCAAA	ATCAAATAA			

60 This encodes a protein having amino acid sequence <SEQ ID 476; ORF48ng-1>:

1 MNIHALLSEQ WTLEPPFLPKR LLLSLILLAPNAVEFWVLAL LTATARPIVN  
 51 LDYLPALLI ALPWRFVKIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI  
 101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTD RHIACAAVV  
 151 AAAGYFTGHL SYDRGRMAN IFGANNFYIA KSQAMLYTVS QNADFITAGL  
 201 VDPVFLPLGN QQRAATRLSE PKSQKILFIV AESWGLPGNP ELQNATFAKL  
 251 LAQKDRFSVW ESGSFPPFIGA TVEGEMRELC AYGGRLRGFAL RRAPDEKFA  
 301 CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFOIKT AENLIGKKT  
 351 AIFGGVCDSE LFGEVSAFFK KHDGGLFYWM TLTSHADYPE SDIFNHLKLC  
 401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIR RPEMKGTEVI IVGDHPPPVG  
 451 NLNETFRYLK QGHVAWLHFK IK\*

ORG48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

		10	20	30	40	50	60
orf48-1.pep		MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVEFWVLALLTATARPIVNL	DYLPALLI				
15	orf48ng-1	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVEFWVLALLTATARPIVNL	DYLPALLI				
		10	20	30	40	50	60
		70	80	90	100	110	120
orf48-1.pep		ALPWRFVKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL					
20	orf48ng-1	ALPWRFVKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf48-1.pep		LLYMLAMPFVLQKAAAKTDFRHIACAAVVAAGYFTGHLSYDRGRMANIFGANNFYIA					
25	orf48ng-1	LLYMLAMPFVLQKAAVKTD	FRHIACAAVVAAGYFTGHLSYDRGRMANIFGANNFYIA				
		130	140	150	160	170	180
		190	200	210	220	230	240
orf48-1.pep		KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSQKILFIVAESWGLPANP					
30	orf48ng-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATRLSEPKSQKILFIVAESWGLPGNP					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf48-1.pep		ELQNATFAKLLAQKDRFSVWESGSFPFIGATVEGEMRELCAYGGRLRGFALRRAPDEKFA					
35	orf48ng-1	ELQNATFAKLLAQKDRFSVWESGSFPFIGATVEGEMRELCAYGGRLRGFALRRAPDEKFA					
		250	260	270	280	290	300
		310	320	330	340	350	360
orf48-1.pep		CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFOEIKTAENLIGKKTCAIFGGVCDSE					
40	orf48ng-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFOIKTAENLIGKKTCAIFGGVCDSE					
		310	320	330	340	350	360
		370	380	390	400	410	420
orf48-1.pep		LFGEVSAFFKKHDKGLFYWMTLTSHADYPESDIFNHLKCTEYGLPAETDLCRNFSLHTQ					
45	orf48ng-1	LFGEVSAFFKKHDKGLFYWMTLTSHADYPESDIFNHLKCTEYGLPAETDLCRNFSLHTQ					
		370	380	390	400	410	420
		430	440	450	460	470	
orf48-1.pep		FFDQLADLIQRPEMKGTEVIIIVGDHPPVGNLNETFRYLKQGHVAWLNFKIKX					
50	orf48ng-1	FFDQLADLIRPEMKGTEVIIIVGDHPPVGNLNETFRYLKQGHVAWLHFKIKX					
		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 57**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

```

      1  ..GTGAGCGGAC GTTACCGCGC TTTGGATCGC GTTTCCAAAA TCATCATCGT
      51  TACTTTGAGT ATCGCCACGC TTGCCGCGCG CGGCATCGCT ATGTCGCGCG
5      101  GTATGCAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGCC
      151  GGTTTGGGCT TCCTGATCGC GCTGATGGGC TGGATGCCCG CGCCGATTGA
      201  AATTTCCGCC ATCAATTCTT TGTGGGTAAC CGAAAAACAA CGCATCAATC
      251  CTTCCGAATA CCGCGACGGG ATTTTGAAT TCAACGTCGG TTATATCGCC
      301  AGTGCGGTTT TGGCTTTGGT TTTCTTGCA CTGGGCGC.G TAGCGCCGAA
10     351  CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAATAT AACGGGCAAT
      401  TGATCAATAT GTACGCC..

```

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

```

      1  ..VSGRYRALDR VSKIIIVTLS IATLAAAGIA MSRGMQMSD FIEPTPWTLA
      51  GLGFLIALMG WMPAPIEISA INSLWVTEKQ RINPSEYRDG IFEFNVGYIA
15     101  SAVLALVFLA LGXVAPNGNG XTVQMAGGKY NGQLINMYA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 479>:

```

      1  ATGTCCGAAC AACATATTTT GACTTGGAAA AGTAAAATCA ACGCATTGGG
      51  TCCGGGGATC ATGATGGCTT CGGCGGCGGT CGGCGGTTTG CACCTGATTG
20     101  CCTCGACGCA GCGGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC
      151  ATCCTGACCA ACCTCTTCAA ATACCCGTTT TTCCGCTTCA GCGCGCATTA
      201  CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
      251  GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT
      301  AACGCGGGCG CGTFCGCCAT TGTAACCGCC GCCATCGTCA AAATGGCGAT
25     351  TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT
      401  CTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
      451  TCCAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCCGCGCG
      501  CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
      551  CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
30     601  ATGCCCCGCG CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
      651  AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTCTA
      701  ACGTCGGTTA TATCGCCAGT GCGGTTTGG CTTTGGTTT CCTGCACTG
      751  GGCGCGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
      801  CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTACC ATCGGCGGCT
      851  GGTCGCGCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCAGC
35     901  ACGATTACCG TCGTGGACGG CTATGCCCGT GCCATTGCCG AACCCGTGCG
      951  COTGCTGCGC GGAAGAGACA AAACGGGCAA CGCCGAATTC TTTGCCTGGA
1001  ATATTTGGGT GCGGGCAGC GGTTTGGCGG TGATTTTCTG GTTGACGGC
1051  GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCCGC
1101  CCCTGTGTTT GCCTGGCTGA ATTACCGTTT GGTAAAGGT GATGAAAAAC
40     1151  ACAAACCTAC ATCAGGTATG AATGCCCTTG CATTGGCAGG CTTGATTTAT
1201  CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGG GAATGTTCAA
1251  ATGA

```

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

```

      1  MSEQHISTWK SKINALGPGI MMASAAVGS HLIASTQAGA LYGWQIALII
45     51  ILTNLFKYPF FRSAHYTLD TGKSLIEGYA EKSRYLWVF LILCILSATI
      101  NAGAVAIITA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV
      151  SKIIIVTISI ATLAAAGIAM SRGMQMSDF IEPTPWTLAG LGFLIALMGW
      201  MPAPIEISAI NSLWVTEKQR INPSEYRDI FEFNVGYIAS AVLALVFLAL
      251  GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT
50     301  TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIIWVAGS GLAVIEWFDG
      351  VMANLLKFAM IAAFVSAPVF AWLNRYLVKG DEKHKLTSGM NALALAGLIY
      401  LTGFTVLFLI NLAGMFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N. meningitidis*:

-283-

```

                    10      20      30
orf53.pep          VSGRYRALDRVSKIIIVTSLIATLAAAGIA
                    |||
orf53a             AAIVKMAIPSLMFDAGTVAALIMASCLIIIVS
5 110      120      130      140      150      160
                    40      50      60      70      80      90
orf53.pep          MSRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAP
                    |||
10 orf53a           MSRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAP
170      180      190      200      210      220
                    100      110      120      130      139
orf53.pep          IFEFNVGYIASAVLALVFLALGXVAPNGXTVQMA
                    ||:|||||:|||||:|||||:|||||:|||||
15 orf53a           IFDFNVGYIASAVLALVFLALGAFVQYNGEAVQMA
230      240      250      260      270      280
                    100      110      120      130      140
orf53a             AFIAFACMYGTTITVVDGYARAIAEPVRLLRGKDK
20 290      300      310      320      330      340

```

The complete length ORF53a nucleotide sequence <SEQ ID 481> is:

```

1  ATGTCCGAAC AACATATTTT GACTTGGAAG AGTAAAATCA ACGCATTGGG
51  ACCGGGGATT ATGATGGCTT CGGCGGCGGT CGGCGGTTTC CACCTGATTG
101 CCTCGACGCA GCGGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC
25 151  ATCCTGACCA ACCTCTTCAA ATACCCGTTT TTCCGCTTCA GCGCGCATT
201  CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
251  GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT
301  AACGCGGGCG CGGTCCGCAT TGTAAACGCC GCCATCGTCA AAATGGCGAT
351  TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT
30 401  CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
451  TCCAAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCCGCCGG
501  CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
551  CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
601  ATGCCGCGCG CGATGCAAT TCCGCCATC AATTCTTTGT GGGTAACCGA
35 651  AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTTC
701  ACGTCGGTTA TATCGCCAGT GCGGTTTTGG CTTTGGTTTT CCTTGCACTG
751  GGCGCGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
801  CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTACC ATCGCGGCT
35 851  GGTCGCGCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCACG
40 901  ACGATTACCG TTGTGGACGG CTATGCCCCT GCCATTGCCG AACCCGTGCG
951  CCTGCTGCGC GGAAGAGACA AAACGGGCAA CGCCGAATTC TTTGCCTGGA
1001 ATATTGGGT GCGGGGAGC GGTGCGCG TGATTTTCTG GTTTGACGGC
1051 GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCCGC
45 1101 CCCTGTGTTT GCCTGGCTGA ATTACCGTTT GGTCAAAGGT GATGAAAAAC
1151 ACAAACTCAC ATCAGGTATG AATGCCCTTG CATGGCAGG CTTGATTTAT
1201 CTGACCGGTT TTACCGTTT GTTCTTATTG AATTGGCGG GAATGTTCAA
1251 ATGA

```

This encodes a protein having amino acid sequence <SEQ ID 482>:

```

1  MSEQHISTWK SKINALGPGI MMASAAVGGG HLIASTQAGA LYGWQIALII
50 51  ILTNLFKYPF FRFSAHYTLT TGKSLIEGYA EKSRYVLWVF LILCILSATI
101 NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV
151 SKIIIVTSLI ATLAAAGIAM SRGMQMOSDF IEPTPWTLAG LGFLIALMGW
201 MPAPIEISAI NSLWVTEKQR INPSEYRDGI FDFNVGYIAS AVLALVFLAL
251 GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT
55 301 TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIIWVAGS GLAVIFWFDG
351 VMANLLKFAM IAAFVSAFVF AWLNRYRLVK DEKHKLTSGM NALALAGLIY
401 LTGFTVLELL NLAGMFK*

```

ORF 53a shows 100.0% identity in 417 aa overlap with ORF53-1:

```

                    10      20      30      40      50      60
60 orf53a.pep      MSEQHISTWKSINALGPGIMMASAAVGGSHLIASTQAGALYGWQIALIIILTNLFKYPF
                    |||
orf53-1            MSEQHISTWKSINALGPGIMMASAAVGGSHLIASTQAGALYGWQIALIIILTNLFKYPF
                    10      20      30      40      50      60
65 70      80      90      100      110      120

```

-284-

	orf53a.pep	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAAIVKMAIPSL	
	orf53-1	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAAIVKMAIPSL	
5		70 80 90 100 110 120	
	orf53a.pep	130 140 150 160 170 180	
10	orf53-1	MFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTLSIATLAAAGIAMSRRGMQMSDF	
		130 140 150 160 170 180	
	orf53a.pep	190 200 210 220 230 240	
15	orf53-1	IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGIFDFNVGYIAS	
		190 200 210 220 230 240	
	orf53a.pep	250 260 270 280 290 300	
20	orf53-1	AVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT	
		250 260 270 280 290 300	
	orf53a.pep	310 320 330 340 350 360	
25	orf53-1	TITVVDGYARAIAEPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM	
		310 320 330 340 350 360	
	orf53a.pep	370 380 390 400 410	
30	orf53-1	IAAFVSAPVFAWLNRYLVKGDEKHKLTSGMNALALAGLIYLTGFTVFLNLAGMFKX	
		370 380 390 400 410	
35			

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF53 shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) from *N. gonorrhoeae*:

40	orf53.pep	VSGRYRALDRVSKIIIVTLSIATLAAAGIA	30
	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTLSIATLAAAGIA	91
	orf53.pep	MSRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	90
45	orf53ng	MSRGMQMOPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	151
	orf53.pep	IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVQMAGGKYNGQLINMYA	139
50	orf53ng	IFDFNVGYIASAVLALVFLALGAFVQYGNGEAVQMGGGKYIGQLINMYAVTIGGGSRPLV	211

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

55	1	MPKKSCVYLW VFLILCIASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
	51	ALIMASCLII IVSGRYRALD RVSKIIIVTL SIATLAAAGI AMSRGMQMOP
	101	DFIEPTPWTL AGLGFLIALM GWMPAPIEIS AINSLWVTEK QRINPSEYRD
	151	GIFDFNVGYI ASAVLALVFL ALGAFVQYGN GEAVQMGGGK YIGQLINMYA
	201	VTIGGGSRPL VAFIAFACMY GAASTVVDGY ARAIAEPVRL LRGKDKTARF
	251	IVLLEKLGR HRFGRDFLV*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

60	1	..aagaAAAGCT GCGTTTATTT GTGGGTTTTT TTGATTTTGT GTATCGCCTC
	51	CGCCACGATT AACGCGGGCG CGGTCGCCAT TGTAAACGCC GCCATCGTCA
	101	AAATGCGGAT TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG



-285-

5  
10  
15

```

151  ATTATGGCAT  CCTGCCTGAT  TATTTTGGTG  AGCGGACGTT  ACCGCGCTTT
201  GGATCGTGTT  TCCAAAATCA  TCATTGTTAC  TTTGAGCATC  GCCACGCTTG
251  CCGCGCGCGG  CATCGCTATG  TCGCGCGGTA  TGCAGATGCA  GCGCGATTTT
301  ATCGAGCCGA  CACCGTGGAC  GCTTGCCGGT  TTGGGCTTCC  TGATCGCGCT
351  GATGGGCTGG  ATGCCCGCGC  CGATCGAAAT  TTCCGCCATC  AATTCCTTGT
401  GGGTAACCGA  AAAACAACGC  ATCAATCCTT  CTGAATACCG  CGACGGGATT
451  TTCGATTTCA  ACGTCGGTTA  TATCGCcagT  GCGGTTTTTG  CTTTGGTTTT
501  CCTTGCACTG  GCGCGGTTTG  TGCAATACGG  CAACGGCGAA  GCAGTGCAGA
551  TGGCGGGCGG  CAAATATATC  GGGCAATGTA  TTAATATGTA  TGCCGTAACC
601  ATCGGCGGCT  GGTCTCGTCC  GCTGGTGGCG  TTTATCGCGT  TTGCCTGTAT
651  GTACGGCACG  ACGATTACCG  TTGTGGACGG  TTATGCGCGT  GCCATTGCCG
701  AACCCTGCGC  CCTGCTGCGC  GGCAGGGATA  AAACCGGCAA  CGCCGAGTTG
751  TTtgccTGGA  ATATTGGGT  GCGGGGACG  GGTTCGGCG  TGATTTCCTG
801  GTTTGACggc  gcaatGGCgG  AActgcTCAA  ATTTGCGATG  ATtgccgcCT
851  TTGTGTCCG  CCCTGTGTT  GCCTGGCTCA  ACTACCGCT  CGTCAAAGGG
901  GACAAACGCC  ACAGGCTTAC  CGCCGGTATG  AACGCCCTTG  CCATTGTGCG
951  CCTGCTCTAC  CTGGCCGGT  TTGCCGTTTT  GTTCCTGTTG  AACCTTACCG
1001 GACTTTTGGC  ATAG

```

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng-1>:

20  
25

```

1  .KKSCVYLWVF LILCIASATI NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL
51  IMASCLIIIV SGRYRALDRV SKIIIVTL SIATLAAAGIAM SRGMQMPDF
101 IEPTPWTLAG LGFLIALMGW MPAPIEISAI NSLWVTEKQR INPSEYRDGI
151 FDFNVGYIAS AVLALVFLAL GAFVQYGNGE AVQMAGGKYI GQLINMYAVT
201 IGWWSRPLVA FIAFACMYGT TITVVDGYAR AIAEPVRLLR GRDKTGNAEL
251 FAWNIVVAGS GLAVIFWFDG AMAELLKFAM IAAFVSAPVF AWLNYRLVKG
301 DKRHRLTAGM NALAIVGLLY LAGFAVLFL NLTGLLA*

```

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

30  
35  
40  
45  
50  
55  
60  
65

```

          60      70      80      90      100     110
orf53-1.pep  ILTNLFKYPFRRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTA
orf53ng-1    :|| ||||| ||||| ||||| |||||
              10      20      30
          120     130     140     150     160     170
orf53-1.pep  AIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTL SIATLAAAGIAM
orf53ng-1    AIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTL SIATLAAAGIAM
              40      50      60      70      80      90
          180     190     200     210     220     230
orf53-1.pep  SRGMQMOSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGI
orf53ng-1    SRGMQMOPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGI
              100     110     120     130     140     150
          240     250     260     270     280     290
orf53-1.pep  FDFNVGYIASAVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGWWSRPLVA
orf53ng-1    FDFNVGYIASAVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGWWSRPLVA
              160     170     180     190     200     210
          300     310     320     330     340     350
orf53-1.pep  FIAFACMYGTTITVVDGYARAIAEPVRLLRGDKDTGNAEFFAWNIVVAGSGLAVIFWFDG
orf53ng-1    FIAFACMYGTTITVVDGYARAIAEPVRLLRGRDKDTGNAELFAWNIVVAGSGLAVIFWFDG
              220     230     240     250     260     270
          360     370     380     390     400     410
orf53-1.pep  VMANLLKFAMIAAFVSAPVFAWLNYRLVKGDKEHKLTSGMNALALAGLIYLTGFTVLFL
orf53ng-1    AMAELLKFAMIAAFVSAPVFAWLNYRLVKGDKRHRLTAGMNALAIVGLLYLAGFAVLFL
              280     290     300     310     320     330
orf53-1.pep  NLAGMFKX
orf53ng-1    NLTGLLAX

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 58

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

```

1  ..TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGTG
51  TGCGCTTGCC GGCTTGT TTTT TGTCCGCGC ACAATCCGAA CGCGAGTGGA
101 TGCGCGAGGT TTCTGCGTGG CAGGAAAAGA AAGGGGAAAA ACAGGCGGAG
151 CTGCCTGAAA TCAAAGACGG TATGCCCGAT TTTCCCGAAC TTGCCCTGAT
201 GCTTTTCCAC GCCGTCAAAA CGGCAGTGTA TTGGCTGTTT GTCGGTGTGC
251 TCCGTTTCTG CCGAACTAT CTGGCGCACG AATCCGAACC GGACAGGCC
301 GTTCCGCCT..

```

This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

```

1  ..LRETAYVLDS FDRYFVVALA GLFFVRAQSE REWMREVSAA QEKKGKQAE
51  LPEIKDGMPE FPELALMLFH AVKTAVYWLF VGVVRFGRNY LAHESEPRP
101  VPP..

```

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

```

20  1  ATGTTTGGGA TAGTTTGTAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51  GTTTTTTGTC CGCGCACAAAT CCGAACCGCA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAATCAAA
151 GACGGTATGC CCGATTTTCC CGAAGTTGCC CTGATGCTTT TCCATGCCGT
201 CAAACCGGCA GTGTATTGGC TGTGTGTCGG TGTCGTCCGT TTCTGCCGAA
25  251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301 GCAAACCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
351 AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
30  501 AATTTGCCCC GTCCGTCGGG TTTTAAAGA AATCACTTTG GAAGAAGCAA
551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
601 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA
651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701 AACGCACGTA TTCCCATATG TTCGATGCGG ACAAGAAGC GTTTTCCGAG
35  751 TCTGCGGATT ACGGATTGTA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCACCGTC
851 ATGCAGGGCA GGGGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC
901 CAAGGGCAGT CCGTTTCAGA CGGCACGGCC GTCCGCGATG CCCGCCGCCG
951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
40  1001 CGCGAATTTT TCGCCTGATT CCGGAAAGTC AGACGGTTGT CGGGAACCGG
1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTCACGG AAACCGTTTC
1101 GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAAACTGCC GATATCCATA
1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
45  1201 CCGAAAGTTC CCATGACCGC AATCGATATT CAGCCGCCGC CTCCCGTATC
1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GTCAGGATTC GAGCAGGTGC
1301 AACGCAGCCG CATTGCCGAG ACCGACCATC TTGCCGATGA TGTTTGAAT
1351 GGAGGTGGC AGGAGGAAC CGCCGTATT GCGGATGACG GCAGTGAGG
1401 TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCGGAAACC GAAGCGTTTCG
1451 GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
50  1501 CCGTCCTGCC GGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCTCATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601 TGCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAACTGTTG
1651 GAAACAGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT
1701 CAAGGTTGTC GATTCTTATT CCGGCCCCGT AATTACCGGT TATGAAATCG
55  1751 AACCCGATGT CGGCGTGGC GGCAATTCCG TTCTGAATCT GGAAGAAAGT
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTGTGCG AAACCATCCC
1851 CGGCAAAACC TGCATGGGTT TGGAACTTCC GAACCGAAA CGCCAAATGA

```

-287-

1901 TACGCCTGAG CGAAATCTTC AATTCGCCCG AGTTTGCCGA ATCCAAATCC  
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC  
 2001 CGACTTGGGA AAAGCACCGC ATTTGTTGGT TGCCGGCAGC ACCGGTTCGG  
 2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC  
 2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAAA TGCTGGAATT  
 2151 GAGCATTAC GAAGGCATCC CGCACCTGCT CGCCCTGTC GTTACCGATA  
 2201 TGAAGCTGGC GGCAACGCG CTGAACCTGGT GTGTTAACGA AATGGAAAAA  
 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGTAATCTTG CGGGCTTCAA  
 2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT  
 2351 TCAGCTCAC GCCCGACGAT CCCGAACCTT TGGAAAACT GCCGTTTATC  
 2401 GTGGTCGTGG TCGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA  
 2451 AATCGAAGAA CTGATTGCCG GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA  
 2501 TCCATTGAT TCTTGCCACA CAACGCCCA GCGTCGATGT CATCAGGGT  
 2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA  
 2601 AATCGACAGC CGCAGATTTC TCGACCAAAT GGGCGCGGAA AACCTGCTCG  
 2651 GTCAGGGCGA TATGCTGTTC CTGCTGCCGG GTACTGCCTA TCCGCAGCGC  
 2701 GTTCACGGCG CGTTTGCTC GGATGAAGAG GTGCACCGCG TGGTCGAATA  
 2751 TTTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATT TTGAGCGGCG  
 2801 GCGGCAGCGA AGAGCTGCCC GGCATCGGGC GCAGCGGCGA CGACGAAACC  
 2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCCTGAAAA CGCGCAAAGC  
 2901 CAGCATTTTC GCGGTACAGC GCGCCTTGGT TATCGGCTAC AACCGCGCCG  
 2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA  
 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGACAATG CTTGA

This corresponds to the amino acid sequence <SEQ ID 490; ORF58-1>:

25 1 MFWIVLVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPEIK  
 51 DGMPDFPELA LMLFHAVKTA VYWLFGVVR FCRNYLAHES EPDRPVPPAS  
 101 ANRADVPTAS DGYSDSGNGT EEAETEEAEA AEEEAADTED IATAVIDNRR  
 151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI  
 201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSHM FDADKEAFSE  
 30 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FHRHAGQKGQ QAEAKSPDVS  
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGKR  
 351 DVEMPSETEN VETETVSSVG YGGPVYDETA DIHIEEPAAP DAWVVEPEEV  
 401 PKVPMTAIDI QPPPPVSEIY NRTYEPSPGF EQVQRSRIAE TDHLADDVLN  
 451 GGWQEETAAI ADDGSEGAAE RSSGQYLSET EAFGHDSQAV CPFENVPSER  
 35 501 PSCRVSDETEA DEGAFPSEET GAVSEHLPTT DLLLPLPLFNP EATQTEEELL  
 551 ENSITIEEKL AEFKVKVKV DSYSGPVITR YEIEPDVGVR GNSVLNLEKD  
 601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS  
 651 KLTILALGQDI TGQPVVTDLG KAPHLLVAGT TSGSKSVGVN AMILSMLFKA  
 701 APEDVRMIMI DPKMLELSIY EGIPHLLAPV VTDMLAANA LNWCVNEMEK  
 40 751 RYRLMSFMGV RNLAGFNQKI AEAARGEKI GNPFSLTPDD PEPLEKLPFI  
 801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG  
 851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LLPGTAYPQR  
 901 VHGAFAFASDEE VHRVVEYLKQ FGEPDYVDDI LSGGSEELP GIGRSGDDET  
 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE  
 45 1001 HGNRTILVP LDNA\*

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF58 shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) from strain A of *N.*

50 *meningitidis*:

orf58.pep      10      20      30      40      50      60  
 LRETAYVLDSFDRYFVVALAGLFFVRAQSEREWMRVSAWQEKKGKEKQAELEPEIKDGMPPD  
 orf58a      MFWIVLVILVILLLALAGLFFVRAQSEREWMRVSAWQEKKGKEKQAELEPEIKDGMPPD  
 55      10      20      30      40      50  
 orf58.pep      70      80      90      100  
 FPTELALMLFHAVKTAVYWLFGVGVVRFRCRNYLAHESEPDRPVPP  
 orf58a      FPTELALMLFHAVKTAVYWLFGVGVVRFRCRNYLAHESEPDRPVPPASANRADVPTASDGYS  
 60      60      70      80      90      100      110

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

```

1  ATGTTTTGGA TAGTTTTGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51  GTTTTTTGTC CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
5  151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201 CAAAACGGCA GTGTATTGGC TGTGTGTCGG TGTCTGCCGT TTCTGCCGAA
251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301 GCAAATCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
10  351 AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451 ATCCCATTCC ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
501 AATTTGCCCC GTCCGTCGGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
601 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA
15  651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701 AACGCACGTA TTCCGATGCG TTCCGATGCG ACAAGAAGC GTTTTCCGAG
751 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TCCGCGCCGC
851 ATGCAGGGCA GGGNAAAGGG CAGGCGGAGG CNAATCCCC GGATGTTTCC
20  901 CAAGGCGAGT CCGTTTCAGA CGGCACAGCC GTCCGCGATG CNGCCGCGC
951 CGTTTCCGTC AATTTGAAGG AACCGAACAA GGCAACGGTT TCTGCGGAGG
1001 CGCGGATTTT GCGCCTGATT CCGGAAAGTC GGACGGTTGT CGGGAAACGG
1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTACAGG AAANTGTTTC
1101 GTCTGTGGGA TACGGCGNTC CGGTTTATGA TGAAACTGCC GATATCCATA
25  1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCAAGC ACCCGAAGTG
1201 CCGAAAGTTC CCATGCCCGC AATNGATATT CCGCCGCCGC CTCCCGTATC
1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GGCAGGATTC GAGCAGGTGC
1301 AACGCAGCCG CATTGCCGAA ACCGATCATC TTGCCGATGA TGTTTTGAAT
1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGAATGACG GCAGTGAGGG
30  1401 TGTGGCAGAG CGGTCAAGCG GGCAATATTT GTCGGAAACC GAAGCGTTCG
1451 GGCATGACAG TCAGGCGGTT TGTCGTTTG AAAATCTGCC GTCTGAACGC
1501 CCGTCCCGCC GGGCATNGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
35  1601 TGCCGCCGCT GTTCAATCCC GGGGCGACGC AAACCGAAGA AGANCTGTTG
1651 GANAACAGCA TCACCATCGA AGAAAAATNG GCGGAGTTCA AAGTCAAGGT
1701 CAAGGTTGTC GATTCTTATT CCGCCCCGT GATTACGCGT TATGAAATCG
1751 AACCCGATGT CGGCGTGCGC GGCAATTCCG TTCTAAATCT GGAAAAAGAN
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCG AAACCATCCT
40  1851 CGGCAAAACC TGTATGGGTT TGGAACCTCC GAACCGGAAA CGCCAAATGA
1901 TACGCTGAG CGAAATCTTC AATTCGCCCG AGTTTGCCGA ATCCAAATCC
1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC
2001 CGACTTGGGC AAAGCACC GC ATTTGTTGGT TGCCGGCACG ACCGGTTCGG
2051 GCAAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
45  2101 GCGCCGGAAG ACGTGCCTAT GATTATGATC GATCCGAAAA TGCTGGAATT
2151 GAGCATTTAC GAAGGCTATC CGCACCTGCT CGCCCCGTGC GTTACCGATA
2201 TGAAGCTGGC GGCAAACGCG CTGAACTGGT GTGTTAACGA AATGGAAAAA
2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CCGGTNTCAA
2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG GGAGAAAATC GGCAACCCGT
2351 TCAGCCTCAC GCCGCAACAT CCGGAACCTT TGGANAAATT GCCGTTTATC
50  2401 GTGGTCGTGG TTGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
2451 AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
2501 TCCATCTTAT CCTTGCCACA CAACGCCCCA GTGTCGATGT CATCACGGGT
2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
55  2601 AATCGACAGC CGCACGATTC TTGACCAAAT GGGTGCGGAA AACCTGCTCG
2651 GGCAGGGCGA TATGCTGTTC CTGCCGCCGG GTACGGCCTA TCCGCAGCGC
2701 GTTCACGGCG CGTTTGCCTC GGATGAAGAG GTGCACCGCG TGGTCGAATA
2751 TTTGAAACAG TTGCGCGAAC CGGACTATGT TGACGATATN TTGAGCGGCG
2801 GTATGTCCGA CGATTTGCTG GGAATCAGCC GGAGCGGCGA CGGCGAAACC
2851 GATCCGATGT ACGACGAGGC CGTGTCNGTT GTTTTGAAAA CGCGCAAAGC
60  2901 CAGCATTCTT GGCGTGCAGC GCGCATTGCG TATCGGCTAT AATCGCGCCG
2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTNGACAATG CTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 492>:

```

1  MFWIVLVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQALPEIK
65  51  DGMPDFPELA LMLFHAVKTA VYWLFVGVVR FCRNYLAHES EPDRPVPPAS
101  ANRADVPTAS DGYSDSGNGT EEAETEEAEA AEEEEADTED IATAVINNR
151  IPFDRSIAEG LMPSESEISP VRPFKEITL EEATRALNSA ALRETKKRYI
201  DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSRM FDADKEAFSE

```

251	SADYGFEPIFY	EKQHPSAFSA	VKAENARNAP	FRRHAGQGKG	QAEAKSPDVS
301	QGQSVSDGTA	VRDAXRRVS	NLKEPNKATV	SAEARISRLI	PESRTVVGKR
351	DVEMPSETEN	VFTXVSSVG	YGXPVYDETA	DIHIEEPAAP	wDAWVVEPEV
401	PKVPMFAXDI	PPPPVSEIY	NRTYEPPAGF	EQVQRSRIAE	TDHLADDVLN
451	GGWQEETAAI	ANDGSEGVAE	RSSGQYLSET	EAFGHDSQAV	CPFENVPSER
501	PSRRAXDTEA	DEGAFQSEET	GAVSEHLPTT	DLLEPLFNP	GATQTEEXLL
551	XNSITIEEKX	AEFKVKVKV	DSYSGPVITR	YEIEPDVGV	GNSVLNLEKX
601	LARSLGVASI	RVVETILGKT	CMGLELPNPK	RQMIRLSEIF	NSPEFAESKS
651	KLTALGQDI	TGQPVVTDLG	KAPHLVAGT	TGSGKSVGVN	AMILSMLFKA
701	APEDVRMIMI	DPKMLELSIY	EGIPHLLAPV	VTDMKLAANA	LNWCVNEMEK
751	RYRLMSFMGV	RNLAGXNOKI	AEAAARGEKI	GNPFLTPDN	PEPLXKLPFI
801	VVVVDEFADL	MMTAGKKIEE	LIARLAQKAR	AAGIHLILAT	QRPSVDVITG
851	LIKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGQGDMLF	LPEGTAYPQR
901	VHGAFASDEE	VHRVVEYLKQ	FGEPTYVDDX	LSGGMSDDL	GISRSGDGET
951	DPMYDEAVSV	VLKTRKASIS	GVQRALRIGY	NRAARLIDQM	EAEGIVSAPE
1001	HNGNRTILVP	XDNA*			

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

		10	20	30	40	50	60
20	orf58a.pep	MFWIVLIVILLLALAGLFFVRAQSEREWREVS	SAWQEKKGKQAE	LPEIKDGM	PDFPELA		
	orf58-1	MFWIVLIVILLLALAGLFFVRAQSEREWREVS	SAWQEKKGKQAE	LPEIKDGM	PDFPELA		
		10	20	30	40	50	60
25	orf58a.pep	LMLFHAVKTAVYWLFVGVVRF	CRNYLAHESE	PDRPVPPAS	ANRADVPTAS	DGYS	DSGNGT
	orf58-1	LMLFHAVKTAVYWLFVGVVRF	CRNYLAHESE	PDRPVPPAS	ANRADVPTAS	DGYS	DSGNGT
		70	80	90	100	110	120
30	orf58a.pep	EEAETEEAEAEAEAA	DTEDIATAVIDN	RRIPFDRSIAE	GLMPSESEIS	PVRPVFKEITL	
	orf58-1	EEAETEEAEAEAEAA	DTEDIATAVIDN	RRIPFDRSIAE	GLMPSESEIS	PVRPVFKEITL	
		130	140	150	160	170	180
35	orf58a.pep	EEATRALNSAALRE	TKRYIDAFEKNET	AVPKVRVSDT	PMGLQIIGLDD	PVLQRTYSRM	
	orf58-1	EEATRALNSAALRE	TKRYIDAFEKNET	AVPKVRVSDT	PMGLQIIGLDD	PVLQRTYSRM	
		190	200	210	220	230	240
40	orf58a.pep	FDADKEAFSESADY	GFEPIFYFEKQHPSAFSA	SAVKAENARNAP	FRRHAGQGKG	QAEAKSPDVS	
	orf58-1	FDADKEAFSESADY	GFEPIFYFEKQHPSAFSA	SAVKAENARNAP	FRRHAGQGKG	QAEAKSPDVS	
		250	260	270	280	290	300
45	orf58a.pep	QGQSVSDGTAVR	DAXRRVSVNLKEPNKATV	SAEARISRLI	PESRTVVGKRD	VEMPSETEN	
	orf58-1	QGQSVSDGTAVR	DAXRRVSVNLKEPNKATV	SAEARISRLI	PESRTVVGKRD	VEMPSETEN	
		310	320	330	340	350	360
50	orf58a.pep	VFTXVSSVGYGXPVY	DETADIHIEEPAAP	DAWVVEPPEV	PKVPMFAXDI	PPPPVSEIY	
	orf58-1	VFTXVSSVGYGXPVY	DETADIHIEEPAAP	DAWVVEPPEV	PKVPMFAXDI	PPPPVSEIY	
		370	380	390	400	410	420
55	orf58a.pep	NRTYEPPAGFEQV	QRSRIAETDHLADD	VLNGGWQEETAAI	ANDGSEGVAE	RSSGQYLSET	
	orf58-1	NRTYEPPAGFEQV	QRSRIAETDHLADD	VLNGGWQEETAAI	ANDGSEGVAE	RSSGQYLSET	
		430	440	450	460	470	480
60	orf58a.pep	EAFGHDSQAVCP	PFENVPSERP	PSRRAXDTEA	DEGAFQSEET	GAVSEHLPTT	DLLEPLFNP
	orf58-1	EAFGHDSQAVCP	PFENVPSERP	PSRRAXDTEA	DEGAFQSEET	GAVSEHLPTT	DLLEPLFNP
		490	500	510	520	530	540
65	orf58a.pep	EEAETEEAEAEAEAA	DTEDIATAVIDN	RRIPFDRSIAE	GLMPSESEIS	PVRPVFKEITL	
	orf58-1	EEAETEEAEAEAEAA	DTEDIATAVIDN	RRIPFDRSIAE	GLMPSESEIS	PVRPVFKEITL	
		490	500	510	520	530	540
70	orf58a.pep	EEAETEEAEAEAEAA	DTEDIATAVIDN	RRIPFDRSIAE	GLMPSESEIS	PVRPVFKEITL	
	orf58-1	EEAETEEAEAEAEAA	DTEDIATAVIDN	RRIPFDRSIAE	GLMPSESEIS	PVRPVFKEITL	
		490	500	510	520	530	540

-290-

5	orf58a.pep	550 560 570 580 590 600	GATQTEEXLLXNSITIEEKXAEFKVKVKVDSYSGPVITRYEIEPDVGVGRNSVLNLEKX
	orf58-1	550 560 570 580 590 600	EATQTEEEELLENSITIEEKLAEFKVKVKVDSYSGPVITRYEIEPDVGVGRNSVLNLEKD
10	orf58a.pep	610 620 630 640 650 660	LARSLGVASIRVVETILGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTLAGQDI
	orf58-1	610 620 630 640 650 660	LARSLGVASIRVVETIPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTLAGQDI
15	orf58a.pep	670 680 690 700 710 720	TGQPVVTDLGKAPHLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY
	orf58-1	670 680 690 700 710 720	TGQPVVTDLGKAPHLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY
20	orf58a.pep	730 740 750 760 770 780	EGIPHLLAPVVTDMLAANALNWCVNEMEKRYRLMSFMGVRNLAGXNQKIAEAAAARGEKI
	orf58-1	730 740 750 760 770 780	EGIPHLLAPVVTDMLAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAAARGEKI
25	orf58a.pep	790 800 810 820 830 840	GNPFSLTPDNPEPLXKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT
	orf58-1	790 800 810 820 830 840	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT
30	orf58a.pep	850 860 870 880 890 900	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQR
	orf58-1	850 860 870 880 890 900	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLLLPGTAYPQR
35	orf58a.pep	910 920 930 940 950 960	VHGAFASDEEVHRVVEYLKQFGEPDYVDDXLSGGMSDDLGLISRSBGDGETDPMYDEAVSV
	orf58-1	910 920 930 940 950 960	VHGAFASDEEVHRVVEYLKQFGEPDYVDDILSGGSEELPGIGRSGDDETDPMYDEAVSV
40	orf58a.pep	970 980 990 1000 1010	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVFXDNAX
	orf58-1	970 980 990 1000 1010	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX

# 50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from *N. gonorrhoeae*:

55	orf58.pep	ALMLEHAVKTAVYWLFVGVVRFRCRNYLAHESEPDRPVPP	103
	orf58ng	SEPDRPVPPASANRADVPTASDGYSDSGNG	30

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

60	1	..SEPDRPVPPA	SANRADVPTA	SDGYSDSGNG	TEEAETEAEE	AAAAEAAADTE
	51	DIATAVIDNR	RIPFDRSIAE	GLMQSESITS	PVRPVFKEIT	LEEATRALSS
	101	AALRETKKRY	IDAFEKNGTA	VPKVRVSDTP	MEGLQIIGLD	DPVLQRTYSR
	151	MFDADKEAFS	ESADYGFEPY	FEKQHPSAFS	AVKAENARNA	PFRRHAGQEK
	201	GQAEAKSPDV	SQGQSVSDGT	AVRDARRRVS	VNLKEPNKAT	VSAEARISRL
	251	IPESRTVVVGK	RDVEMPSETE	NVFTETVSSV	GYGGFPVYDEA	ADIHIEEPAA
	301	PDAWVVEPPE	VPEVAVPEID	ILPPPPVSEI	YNRTYEPPAG	FEQAQRSRIA

-291-

351	ETDHLAADVL	NGGWQEETAA	IADGSEGA	ERSSGQYLSE	TEAFGHDSQA
401	VCPFEDVPSE	RPSCRVSDE	ADEGAFQSEE	TGAVSEHLPT	TDLLPLPLEN
451	PEATQTEEL	LENSITIEEK	LAEFKVKVKV	VDSYSGPVIT	RYEIEPDVGV
501	RGNSVLNLEK	DLARSLGVAS	IRVVETIPGK	TCMGLELNP	KRQMIRLSEI
551	FNSPEFAESK	SKLTALGQD	ITGQPVVTDL	GKAPHLVAG	<u>TTGSGKSVGV</u>
601	NAMILSMLFK	AAPEDVRMIM	IDPKMLELSI	YEGITHLLAP	VVTDMKLAAN
651	ALNWCVNEME	KRYRLMSFMG	VRNLAGFNQK	IAEAAARGEK	IGNPFSLTPD
701	DPEPLEKLPF	IVVVDEFAD	LMMTAGKKIE	ELIARLAQKA	RAAGIHLILA
751	TQRPSVDVIT	GLIKANIPTR	IAFQVSSKID	SRTILDQMG	ENLLGQGDML
801	FLPPGTAYPQ	RVHGAFASDE	EVHRVVEYLK	QFGEPTYVDD	ILSGGSGSEEL
851	PGIGRSGDGE	TDFMYDEAVS	VVLKTRKASI	SGVQRALRIG	YNRAARLIDQ
901	MEAEIVSAP	EHNGNRTILV	PLDNA*		

This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK (accession number p46889) show a 65 % amino acid identity in 459 overlap:

ORF58ng:	467	IEEKLAEFKVKVKVDSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET	526
		+E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE	
FtsK:	868	VEARLADRIKADVNNYSPGPVITRFELNLAGVKAARISNLSRDLARSLSTVAVRVVEV	927
ORF58ng:	527	IPGKTCMGLELNPKNRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL	586
		IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL	
FtsK:	928	IPGKPYVGLELNPKNRQTVYLREVLDAKFRDNPSPLTVVLGKDIAGEPVVADLAKMPHL	987
ORF58ng:	587	LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK	646
		LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK	
FtsK:	988	LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLEVTDMK	1047
ORF58ng:	647	LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTDDPEP--	704
		AAANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +	
FtsK:	1048	DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPPIPDYWKPGDSMDAQH	1107
ORF58ng:	705	--LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL	762
		L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL	
FtsK:	1108	PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL	1167
ORF58ng:	763	IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQRVHGAFASDEEV	822
		IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV	
FtsK:	1168	IKANIPTRIAFTVSSKIDSRTILDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV	1227
ORF58ng:	823	HRVVEYLKQFGEPTYVDDILSGGSGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG	882
		H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG	
FtsK:	1228	HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG	1286
ORF58ng:	883	VQRALRIGYNRAARLIDQMEAEIVSAPAHNGNRTILVP	921
		VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P	
FtsK:	1287	VQRQFRIGYNRAARIIEQMEAQGVISEQGHNGNREVLAP	1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

1	ATGTTTTGGA	TAGTTTTGAT	CGTTATtgtg	TTGCTTGCGC	TTGCCGGCCT
51	GTTTTTTGTC	CGCGCACAAAT	CCGAACGCGA	GTGGATGCGC	GAGGTTTCTG
101	CGTGGCAGGA	AAAGAAAGGG	GAAAAACAGG	CGGAGCTGCC	TGAAATCAAA
151	GACGGTATGC	CCGATTTTCC	CGAGTTTTC	CTGATGCTTT	TCCATGCCGT
201	CAAAACGGCA	GTGTATTGGC	TGTTTGTCGG	TGTCGTCCGT	TTCTGCCGAA
251	ACTATCTGGC	GCACGAATCC	GAACCGGACA	GGCCCGTTCC	GCCTGCTTCT
301	GCAAACCGTG	CGGATGTTCC	GACCGCATCC	GACGGGTATT	CAGACAGTGG
351	AAACGGGACG	GAAGAAGCGG	AAACGGAAGC	AGCAGAAGCT	GCGGAGGAAG
401	AGGCTGCCgA	TACgGAAGAC	ATTGCAACTG	CCGTAATCGA	CAACCGCCGC
451	ATCCcatTCG	ACCGGAGTAT	TGCTGAAGGG	TTGATGCAGT	CTGAAAGCAA
501	AACTTCGCCC	GTCCGTCCGG	TTTTTAAGGA	AATCACTTTG	GAAGAAGCAA
551	CGCGTGCTTT	AAGCAGCGCG	GCTTTAAGGG	AAACGAAAAA	ACGCTATATC
601	GATGCATTTG	AGAAAAACGG	AACAGCCGTC	CCCAAAGTAC	GCGTGTCCGA
651	TACCCCGATG	GAAGGGCTGC	AGATTATCGG	TTTGGACGAC	CCTGTGCTTC
701	AACGCACGTA	TTCCCGTATG	TTTGATGCGG	ACAAAGAAGC	GTTTTCGAG
751	TCTGCGGATT	ACGGATTTGA	GCCGTATTTT	GAGAAGCAGC	ATCCGTCTGC

	801	CTTTTCTGCA	GTCAAAGCCG	AAAATGCACG	GAATGCGCCG	TTCCGCCGTC
	851	ATGCAGGGCA	GGAGAAAGGG	CAGGCGGAGG	CAAAATCCCC	GGATGTTTCC
	901	CAAGGCGAGT	CCGTTTCAGA	CGGCACAGCC	GTCCGCGATG	CCCGCCGCCG
5	951	CGTTTCCGTC	AATTTGAAAG	AACCGAACAA	GGCAACGGTT	TCTGCGGAGG
	1001	CGCGGATTTC	GCGCCTGATT	CCGGAAAGTC	GGACGGTTGT	CGGGAAACGG
	1051	GATGTCGAAA	TGCCGTCTGA	AACCGAAAAT	GTTCACCG	AAACCGTTTC
	1101	GTCTGTGGGA	TACGGCGGTC	CGGTTTATGA	TGAAGCTGCC	GATATCCATA
	1151	TTGAAGAGCC	TGCCGCGCCC	GATGCTTGGG	TGGTCGAACC	ACCCGAAGTG
10	1201	CCGGAGGTAG	CCGTACCCGA	AATCGATATT	CTGCCGCCGC	CTCCCGTATC
	1251	GGAAATCTAC	AACCGTACCT	ATGAGCCGCC	GGCAGGATTG	GAGCAGGCGC
	1301	AACGCGAGCG	CATTGCCGAA	ACCGACCATC	TTGCCGCTGA	TGTTTTGAAT
	1351	GGAGGTTGGC	AGGAGGAAAC	CGCCGCTATT	GCAGATGACG	GCAGTGAGGG
	1401	TGCGGCAGAG	CGGTCAAGCG	GGCAATATCT	GTGGAAGACC	GAAGCGTTTC
	1451	GGCATGACAG	TCAGGCGGTT	TGTCCGTTTG	AAGATGTGCC	GTCTGAACGC
15	1501	CGTCCTGCC	GGGTCTCGGA	TACGGAAGCG	GATGAAGGGG	CGTTCCAATC
	1551	GGAAGAGACC	GGTGCGGTAT	CCGAACACCT	GCCGACAACC	GACCTGCTTC
	1601	TGCCTCCGCT	GTTCAATCCC	GAGGCGACGC	AAACCGAAGA	AGAACTGTTG
	1651	GAAGACAGCA	TCACCATCGA	AGAAAAATG	GCGGAGTTCA	AAGTCAAGGT
20	1701	CAAGGTTGTC	GATTCCTATT	CCGCCCCCGT	GATTACGCGT	TATGAAATCG
	1751	AACCCGATGT	CGGCGTCCGC	GGCAATTCCG	TTCTGAATTT	GGAAAAAGAC
	1801	TTGGCGCGTT	CGCTCGCGCT	GGCTTCCATC	CGCGTTGTCT	AAACCATCCC
	1851	CGGCAAAACC	TGCATGGGTT	TGGAACCTCC	GAACCCGAAA	CGCCAAATGA
	1901	TACGCTGAG	CGAAATTTTC	AATTCGCCCG	AGTTTGCCGA	ATCCAAATCC
	1951	AAGCTGACGC	TCGCGCTCGG	TCAGGACATT	ACCGGACAGC	CCGTCTTAAC
25	2001	CGACTTGGGC	AAAGCACCGC	ATTGCTGGT	TGCCGCGCAG	ACCGGTTCCG
	2051	GCAATCGGT	GGGTGTCAAC	GCGATGATTC	TGTCTATGCT	TTTCAAAGCC
	2101	GCGCCGGAAG	ACGTGCGTAT	GATTATGATC	GATCCGAAAA	TGCTGGAATT
	2151	GAGCATTTAC	GAAGGCATCA	CGCACCTGCT	CGCCCTGTCT	GTTACCGATA
30	2201	TGAAGCTGGC	GGCAAACGCG	CTGAACTGGT	GTGTTAACGA	AATGGAAAAA
	2251	CGCTACCGCC	TGATGAGCTT	TATGGGCGTG	CGCAATCTTG	CGGGCTTCAA
	2301	CCAAAAATC	GCCGAAGCCG	CAGCAAGGGG	AGAAAAATC	GGCAATCCGT
	2351	TACGCTCAC	GCCCGACGAT	CCCGAACCTT	TGGAAAAACT	GCCGTTTATC
	2401	GTGCTCGTGG	TCGATGAGTT	TGCCGATTTC	ATGATGACGG	CAGGCAAGAA
35	2451	AATCGAAGAA	CTGATTGCGC	GCCTCGCCCA	AAAAGCCCGC	GCGGCAGGCA
	2501	TCCACCTTAT	CCTTGCCACA	CAACGCCCCA	GCGTCGATGT	CATCACGGGT
	2551	CTGATTAAGG	CGAACATCCC	GACGCGTATC	GCGTTCCAAG	TGTCCAGCAA
	2601	AATCGACAGC	CGCACGATTC	TCGACCAAAT	GGGCGCGGAA	AACCTGCTCG
	2651	GTCAGGGCGA	TATGCTGTTT	CTGCCGCCCG	GTACTGCCTA	TCCGCAGCGC
40	2701	GTTACGGCG	CGTTTGCCCT	GGATGAAGAG	GTGCACCGCG	TGGTCGAATA
	2751	TCTGAAGCAG	TTTGCCGAGC	CGGACTATGT	TGACGATATT	TTGAGCGCGC
	2801	GCGGCAGCGA	AGAGCTGCCC	GGCATCGGGC	GCAGCGGCGA	CGGCGAAACC
	2851	GATCCGATGT	ACGACGAGGC	CGTATCCGTT	GTCTGAAAAA	CGCGCAAAGC
	2901	CAGCATTTTC	GGCGTACAGC	GCGCCTTGCG	CATCGGCTAC	AACCGCGCCG
	2951	CGCGTCTGAT	TGACCAATG	GAAGCGGAAG	GCATTGTGTC	CGCACCGGAA
45	3001	CACAACGGCA	ACCGTACGAT	TCTCGTCCCC	TTGGACAATG	CTTGA

This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

	1	MEWIVLVIV	LLALAGLEFFV	RAQSEREWMR	EVSAAQEKKG	EKQAELEIK
	51	DGMPDFPEFS	LMLFHAVKTA	VYWLFGVVR	FCRNYLAHES	EPDRPVPPAS
50	101	ANRADVPTAS	DGYSDSGNGT	EEAEETAAEA	AEEEAADTED	IATAVIDNRR
	151	IPFDRSIAEG	LMQSEKSTSP	VRPVFKEITL	EEATRALSSA	ALRETKKRYI
	201	DAFEKNGTAV	PKVRVSDTFM	EGLQIIGLDD	PVLQRTYSRM	FDADKEAFSE
	251	SADYGFEPYF	EKQHPSAFSA	VKAENARNAP	FRRHAGQEK	QAEAKSPDVS
	301	QGQSVSDGTA	VRDARRRVSV	NLKEPNKATV	SAEARISRLI	PESRTVVGKR
55	351	DVEMPSETEN	VFTETVSSVG	YGGFPVDEAA	DIHIEEPAAP	DAWVVEPPEV
	401	PEVAVPEIDI	LPPPPVSEIY	NRTYEPPAGF	EQAQRSRIAE	TDHLAADVLN
	451	GGWQEETAAI	ADDGSEGAEE	RSSGQYLSET	EAFGHDSQAV	CPFEDVPSE
	501	PSCRVSDETEA	DEGAFQSEET	GAVSEHLPTT	DLLEPLFNP	EATQTEELL
	551	ENSITIEEKL	AEFKVKVKVV	DSYSGPVITR	YEIEPDVGV	GNSVLNLEKD
60	601	LARSLGVASI	RVVETIPGKT	CMGLELPNPK	RQMIRLSEIF	NSPEFAESKS
	651	KLTLALGQDI	TGQPVVTDLG	KAPHLVAGT	TGSGKSVGVN	AMILSMLEKA
	701	APEDVRMIMI	DPKMLELSTY	EGITHLLAPV	VTDMKLAANA	LNWCVNEMEK
	751	RYRLMSFMGV	RNLAFNQKI	AEAAARGEKI	GNPFSLTDD	PEPLEKLEFI
	801	VYVVEFADL	MMTAGKIEE	LIARLAQKAR	AAGIHLILAT	QRPSVDVITG
65	851	LIKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGQGDMLF	LPPGTAYPQR
	901	VHGAFADEE	VHRVVEYLKQ	FGEPTYVDDI	LSGGGSEELP	GIGRSGDGET
	951	DPMYDEAVSV	VLKTRKASIS	GVQRALRIGY	NRAARLIDQM	EAEGIVSAPE
	1001	HNGNRTILVP	LDNA*			

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:



-293-

		10	20	30	40	50	60
	orf58-1.pep	MFWIVLIVILLALAGLFFVRAQSEREW	MREVS	SAWQEKKG	QKQAE	LP	IKDGM
	orf58ng-1	MFWIVLIVIVLLALAGLFFVRAQSEREW	MREVS	SAWQEKKG	QKQAE	LP	IKDGM
5		10	20	30	40	50	60
	orf58-1.pep	LMLFHAVKTAVYWL	FVGVVRF	CRNYLA	HESE	PDRPV	PPASAN
	orf58ng-1	LMLFHAVKTAVYWL	FVGVVRF	CRNYLA	HESE	PDRPV	PPASAN
10		70	80	90	100	110	120
	orf58-1.pep	LMLFHAVKTAVYWL	FVGVVRF	CRNYLA	HESE	PDRPV	PPASAN
	orf58ng-1	LMLFHAVKTAVYWL	FVGVVRF	CRNYLA	HESE	PDRPV	PPASAN
		70	80	90	100	110	120
	orf58-1.pep	EEAETEEAEAEAE	EAADTE	DIATA	VIDN	RRIP	PFDRS
	orf58ng-1	EEAETEEAEAEAE	EAADTE	DIATA	VIDN	RRIP	PFDRS
15		130	140	150	160	170	180
	orf58-1.pep	EEAETEEAEAEAE	EAADTE	DIATA	VIDN	RRIP	PFDRS
	orf58ng-1	EEAETEEAEAEAE	EAADTE	DIATA	VIDN	RRIP	PFDRS
		130	140	150	160	170	180
	orf58-1.pep	EEATRALNSAAL	RETKRY	IDAFE	KNET	AVPK	VRVSD
	orf58ng-1	EEATRALSSAAL	RETKRY	IDAFE	KNGT	AVPK	VRVSD
20		190	200	210	220	230	240
	orf58-1.pep	EEATRALNSAAL	RETKRY	IDAFE	KNET	AVPK	VRVSD
	orf58ng-1	EEATRALSSAAL	RETKRY	IDAFE	KNGT	AVPK	VRVSD
		190	200	210	220	230	240
	orf58-1.pep	FDADKEAFSE	SADYGF	EPYFE	KQHPS	AFSA	VAEN
	orf58ng-1	FDADKEAFSE	SADYGF	EPYFE	KQHPS	AFSA	VAEN
25		250	260	270	280	290	300
	orf58-1.pep	FDADKEAFSE	SADYGF	EPYFE	KQHPS	AFSA	VAEN
	orf58ng-1	FDADKEAFSE	SADYGF	EPYFE	KQHPS	AFSA	VAEN
		250	260	270	280	290	300
	orf58-1.pep	QGQSVSDGT	AVRDAR	RRVSV	NLKE	PNKAT	VSAE
	orf58ng-1	QGQSVSDGT	AVRDAR	RRVSV	NLKE	PNKAT	VSAE
30		310	320	330	340	350	360
	orf58-1.pep	QGQSVSDGT	AVRDAR	RRVSV	NLKE	PNKAT	VSAE
	orf58ng-1	QGQSVSDGT	AVRDAR	RRVSV	NLKE	PNKAT	VSAE
		310	320	330	340	350	360
	orf58-1.pep	VFTETVSSV	GGPVY	DETAD	IHIEE	PAPDA	WVVE
	orf58ng-1	VFTETVSSV	GGPVY	DEAADI	HIEE	PAPDA	WVVE
35		370	380	390	400	410	420
	orf58-1.pep	VFTETVSSV	GGPVY	DETAD	IHIEE	PAPDA	WVVE
	orf58ng-1	VFTETVSSV	GGPVY	DEAADI	HIEE	PAPDA	WVVE
		370	380	390	400	410	420
	orf58-1.pep	NRTYEP	PPSGFE	QVQSR	IAETD	HLADD	VLNG
	orf58ng-1	NRTYEP	PAGFEQ	AQSR	IAETD	HLAAD	VLNG
40		430	440	450	460	470	480
	orf58-1.pep	NRTYEP	PPSGFE	QVQSR	IAETD	HLADD	VLNG
	orf58ng-1	NRTYEP	PAGFEQ	AQSR	IAETD	HLAAD	VLNG
		430	440	450	460	470	480
	orf58-1.pep	EAFGHDSQ	AVCPF	ENVP	SERP	SCRVS	DTAE
	orf58ng-1	EAFGHDSQ	AVCPF	EDVP	SERP	SCRVS	DTAE
45		490	500	510	520	530	540
	orf58-1.pep	EAFGHDSQ	AVCPF	ENVP	SERP	SCRVS	DTAE
	orf58ng-1	EAFGHDSQ	AVCPF	EDVP	SERP	SCRVS	DTAE
		490	500	510	520	530	540
	orf58-1.pep	EATQTEEE	LLENS	ITIEE	KLAE	FKVK	VVDS
	orf58ng-1	EATQTEEE	LLENS	ITIEE	KLAE	FKVK	VVDS
50		550	560	570	580	590	600
	orf58-1.pep	EATQTEEE	LLENS	ITIEE	KLAE	FKVK	VVDS
	orf58ng-1	EATQTEEE	LLENS	ITIEE	KLAE	FKVK	VVDS
		550	560	570	580	590	600
	orf58-1.pep	LARSLG	VASIR	VVETI	PGKTC	MGLE	LPNPK
	orf58ng-1	LARSLG	VASIR	VVETI	PGKTC	MGLE	LPNPK
55		610	620	630	640	650	660
	orf58-1.pep	LARSLG	VASIR	VVETI	PGKTC	MGLE	LPNPK
	orf58ng-1	LARSLG	VASIR	VVETI	PGKTC	MGLE	LPNPK
		610	620	630	640	650	660
	orf58-1.pep	TGQP	VVTD	LGKAP	HLVAG	TTG	SGKSV
	orf58ng-1	TGQP	VVTD	LGKAP	HLVAG	TTG	SGKSV
60		670	680	690	700	710	720
	orf58-1.pep	TGQP	VVTD	LGKAP	HLVAG	TTG	SGKSV
	orf58ng-1	TGQP	VVTD	LGKAP	HLVAG	TTG	SGKSV
		670	680	690	700	710	720

		730	740	750	760	770	780
	orf58-1.pep	EGIPHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
5	orf58ng-1	EGITHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58-1.pep	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
10	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLQGQDMLFLLPGTAYPQR					
15	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLQGQDMLFLLPGTAYPQR					
		850	860	870	880	890	900
	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGEFDPYVDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
20	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGEFDPYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX					
25	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	

30 Furthermore, ORF58ng-1 shows significant homology to the *E.coli* protein FtsK:

	sp P46889 FTSK_ECOLI CELL DIVISION PROTEIN FTSK >gi 1651412 gnl PID d1015290 (D1 division protein FtsK [Escherichia coli] >gi 1651418 gnl PID d1015296 (D90727) Cell division protein FtsK [Escherichia coli] >gi 1787117 (AE000191) cell division protein FtsK [Escherichia coli] Length = 1329
35	Score = 576 bits (1469), Expect = e-163 Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)
	Query: 556 IEEKLAEFKVKVKKVVDSSYSGPVITRYEIEPDVGVRGNSVLNLEKDLARSLGVASIRVVET 615 +E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE
40	Sbjct: 868 VEARLADFRIKADVNNYSPGPVITRFEFLNLAGVGKAAARISNLSRDLARSLSTVAVRVVEV 927
	Query: 616 IPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLKGAPHL 675 IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL
45	Sbjct: 928 IPGKPYVGLELPNKKRQTVYLVREVLDNAKFRDNPSPLTVVLGKDIAGEFVVADLAKMPHL 987
	Query: 676 LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK 735 LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK
	Sbjct: 988 LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLTEVVVTDMK 1047
50	Query: 736 LAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP-- 793 AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +
	Sbjct: 1048 DAANALRWCVNEMERRYKLSALGVRNLAGYNEKIAEADRMMPIDPDYWKPGDSMDAQH 1107
55	Query: 794 --LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL 851 L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL
	Sbjct: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL 1167
	Query: 852 IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLQGQDMLFLLPGTAYPQRVHGAFASDEEV 911 IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV
60	Sbjct: 1168 IKANIPTRIAFTVSSKIDSRTILDQAGAESLLGMGDMLYSGPNSTLPPVRVHGAFVRDQEV 1227
	Query: 912 HRVVEYLKQFGEFDPYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG 971 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG
65	Sbjct: 1228 HAVVQDWKARGRPQYVVGITSDSESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG 1286
	Query: 972 VQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVP 1010 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P
	Sbjct: 1287 VQRQFRIGYNRAARIIEQMEAQGIVSEQGHNGNREVLAP 1325

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 59

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:

```

5      1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
      51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
      101 TGCTCGGCCG TGCCGCCGAC GGCC..GTGA TCGCCATCGA TGCCGTGTTG
      151 GCATTGGTCG GCTTCTGGGT C.....
//
10     901  ....A TTGCCATCGG TTTGTTTTTA ATTTACCAA ACAGGCTGAC
      951  CCTGCTTTTT GAAGCCGTGG AAGACGGCAA AATCCATTT TGGCTCGGAC
     1001  TGCTGCCTAT GCACATTATC ATGTTTGTCC TTGCACTCAT CCTGTTGCGC
     1051  GTCCGCAGTA TGCCAGCCA GCCCTTCTGG CAGGCGGTTG GCAAAAGTCT
     1101  GACATTGAAA GCGGAAAAAT GA

```

15 This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GXVIAIDAVL
      51  ALVGFWV... ..
//
     301  ...IAIGLFL IYQNLTLFF EAVEDGKIHF WLGLLPMHII MFVLALILLR
     351  VRSMPSQPFW QAVGKSLTLK GK*

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

      1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
      51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
     101  TGCTCGGCCG TGCCGCCGAC GGCGGTGTCG CCATCGATGC CGTGTGGCA
     151  TTGGTCGGCT TCTGGGTGTC CGGTATGACG CCGCTTTTGC TGGTGTGAC
     201  CGCATTTATC AGTACGTTGA CCGTGTGAC CCGCTACTGG CGCGACAGCG
     251  AAATGTCGGT CTGGCTATCC TGCGGATTGG CATTGAAACA ATGGATACGC
     301  CCGGTGATGC AGTTTGCCGT GCGGTTTGCC GTTTTGGTTG CCGTCATGCA
     351  GCTTGGGTG ATACGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
     401  TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGAGTTCAAC
     451  AGTTTGGGCA AGCGCAACG CAGGGTTTAT TTTGTGCGAA CCTTCGATAC
     501  CGAATCCGGC ATCATGAAAA ACCTGTTCCT GCGCGAACAG GACAAAAACG
     551  CCGCGGACAA CATCATCTTC GCCAAAGAAG GTAACCTCTC GCTGAACGAC
     601  AACAAACGCA CGCTCGAATT GCGCCACGGC TACCGTTACA GCGGCACGCC
     651  CGGACGCGCC GACTACAATC AGGTTTCCTT CCAAAACTC AACCTGATTA
     701  TCAGACACCAC GCCCAAACCT ATCGACCCCG TTTCACCCG CCGTACCATT
     751  CCGACCGCCC AACTGATTGG CAGCAGCAAC CCGCAACATC AGGCGGAATT
     801  GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCCTGCTTG
     851  CCGTGCCGCT TTCTATTTC AACCCGCGCA GCGACATAC CTACAATATC
     901  TTGATTGCCA TCGGTTTGTT TTTAATTAC CAAAACGGGC TGACCCTGCT
     951  TTTTGAAGCC GTGGAAGACG GCAAATCCA TTTTGGGCTC GGACTGCTGC
    1001  CTATGCACAT TATCATGTTT GCCGTTGCAC TCATCCTGTT GCGCGTCCGC
    1051  AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
    1101  GAAAGGCGGA AAATGA

```

45 This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
      51  LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
     101  PVMQFAVPFA VLVAVMQLWV IPWAE LRSRE YAEILKQKE LSLVEAGEFN
     151  SLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNIIF AKEGNFSLND
     201  NKRTLRLRHG YRYSGPGRG DYNQVSFQKL NLIISTTPKL IDPVSHRRTI
     251  PTAQLIGSSN PQHQAEMLWR ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
     301  LIAIGLFLIY QNLTLTLEA VEDGKIHFVW GLLPMHII MF AVALILLRVR
     351  SMPSQPFWQA VGKSLTLKGG K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) from strain A of *N. meningitidis*:

5	orf101.pep	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVWX	10	20	30	40	50
	orf101a	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGXAADXRX-AIDAVLALVGFVWXXM	10	20	30	40	50
10	orf101.pep	.....IAIGLFLIYQNGLTLLFEAVEDGKIHFVWLGL	90	100	110		
	orf101a	LTVSVLLCLLAVPLSYFNPRSGHTYINILXAIGLFLIYQNGLTLLFEAVEDGKIHFVWLGL	280	290	300	310	320
15	orf101.pep	LPMHIIMFVLALILLRVRSMPSQPFVQAVGKSLTLKGGKX	120	130	140	150	
	orf101a	LPMHIIMFVIAIVLLRVRSMPSQPFVQAVGKSLTLKGGKX	340	350	360	370	

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

1	ATGATTTATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CCGCCGTCGG
51	CATTTTCGTC	GTCCTCTTGG	CGGTATTGGT	CTCCACGCAG	GCAATCAACC
101	TGCTCGGCCN	TGCCGCGGAC	NGGCGTNTCG	CCATCGATGC	CGTGTGGCA
25	TTGGTCGGCT	TCTGGGTCNN	NNGNATGACG	CCGCTTTTGC	TNGTGTGAC
201	CGCATTTATC	AGTACGTTGA	CCGTGTTGAC	CCGCTACTGG	CGNGACAGCG
251	AAATGTGGGT	CTGGNTATCC	TGCGGATTGG	CATTGAAACA	ATGGATACGC
301	CCGGTGATGC	AGTTTGCCGT	GCCGTTTGCC	GTTTGTGGTG	CCGTCATGCA
351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTACG	CAGCCGCGAA	TACGCTGAAA
401	TCCTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGGAGGCAGG	CGGGTTC AAC
451	AGTTTGGGCA	AGCGCAACGG	CAGGGTTTAT	TTTGTGCAAA	CCTTCGATAC
501	CGAATCCGGC	ATCATGAAAA	ACCTGTTCTT	GCGCGAACAG	GACAAAAACG
551	CGGCGGACAA	CATCATCTTC	NCCAAAGAAA	GTAACCTTCT	GCTGAACGAC
601	AACAAACGCA	CGCTCGAATT	GCGCCACGGC	TACCGTTACA	GCGGCACGCG
35	CGGACGCGCC	GACTACAATC	AGGTTTCCTT	CCNAAAACCT	AACCTGATTA
701	TCAGACCAC	GCCCAAACCT	ATCGACCCCG	TTTCCCACCG	CCGTACNATN
751	CCNACNGCCC	AACTGATTGG	CAGCAGCAAC	CCGCAACATC	ANGCGGAATT
801	ATGTGGGCGC	ATCTCGCTGA	CCGTGAGCGT	CCTCTACTCT	TGCTGCTTGT
851	CCGTGCGGCT	TTCCTATTTT	AACCCGCGCA	GCGGACATAC	CTACAATATC
40	TTGANTGCCA	TGGTTTGTGT	TTAATTATAC	CAAACGCGGC	TGACCTGTCT
951	TTTTGAAGCC	GTGGAAGACG	GCAAAATCCA	TTTTTGCTCT	GGACTGTCTG
1001	CTATGCACAT	CATCATGTTT	GTCATCGCAA	TGCTACTTCT	GCGCGTCCGC
1051	AGCATGCCCA	GCCAGCCCTT	CTGGCAGGCG	GTGGCAAAA	GTCTGACATT
1101	GAAAGCGGGA	AAATGA			

45 This encodes a protein having amino acid sequence <SEQ ID 502>:

1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGXAAD	XRXAIDAVLA
51	LVGFWVXXMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWXS	CGLALKQWIR
101	PVMQFAVPFA	VLVAVMQLWV	IPWAE LRSRE	YAEILKQKQE	LSLVEAGGFN
151	SLGKRNGRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIF	XKESNFSLND
50	NKRTLELRHG	YRYSGTPGRA	DYNQVSFXKL	NLIISTTPKL	IDPVSHRRTX
251	PTAQLIGSSN	POHXAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
301	LXAIGLFLIY	QNGLTLLFEA	VEDGKIHFVWL	GLLPMHIIMF	VIAIVLLRVR
351	SMPSQPFVQA	VGKSLTLKGG	K*		

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

55	orf101a.pep	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGXAADXRXAIDAVLALVGFVWVXXMT	60
	orf101-1	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWVIGMT	60
60	orf101a.pep	PLLLVLTAFI STLTVLTRYWRDSEMSVWXS CGLALKQWIRPVMQFAVPFAVLVAVMQLWV	120
	orf101-1	PLLLVLTAFI STLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAVLVAVMQLWV	120

5	orf101a.pep	IPWAE LRSREYAEILKQKQELSLVEAGGFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ 180
	orf101-1	IPWAE LRSREYAEILKQKQELSLVEAGEFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ 180
10	orf101a.pep	DKNGGDNIIIFXKESNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFXKLNIISTTPKL 240
	orf101-1	DKNGGDNIIIFAKEGNFSLNDNKRTLELRHGYRYSGTPGRADYNQVSFQKLNIISTTPKL 240
15	orf101a.pep	IDPVSHRRTXPTAQLIGSSNPQHXAELMWRLSLTVSVLLCLLAVPLSYFNPRSGHTYNI 300
	orf101-1	IDPVSHRRTIPTAQLIGSSNPQHQAELMWRLSLTVSVLLCLLAVPLSYFNPRSGHTYNI 300
20	orf101a.pep	LXAIGLFLIYQNGLTLLFEAVEDGKIHFGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA 360
	orf101-1	LIAIGLFLIYQNGLTLLFEAVEDGKIHFGLLPMHIIMFAVALILLRVRSMPSQPFWQA 360
	orf101a.pep	VGKSLTLKGGK 371
	orf101-1	VGKSLTLKGGK 371

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N.*

### *gonorrhoeae*:

30	orf101.pep	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVW 57
	orf101ng	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRV-AIDAVLALVGFVWIGM 59
35	orf101.pep	IAIGLFLIYQNGLTLLFEAVEDGKIHFGLG 333
	orf101ng	SLTVSVLLCLLAVPLSYFNPRSGHTYNI LIAIGLFLIYQNGLTLLFEAVEDGKIHFGLG 331
	orf101.pep	LLPMHIIMFVLALILLRVRSMPSQPFWQAVGKSLTLKGGK 373
	orf101ng	LLPMHIIMFVIAIVLLRVRSMPSQPFWQAVG 362

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial amino acid sequence <SEQ ID 504>:

45	1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGRAAD	GRVAIDAVLA
	51	LVGEFVIGMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWLS	CGLALKQWIR
	101	PVMQFAVPFA	ILIAVMQLWV	IPWAE LRSRE	YAEILKQKQE	LSLVEAGEFN
	151	NLGKRNGRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIIF	AKEGNFSLKD
50	201	NKRTLELRHG	YRYSCTPGRA	DYNQVSFQKL	NLIISTTPKL	IDPVSHRRTI
	251	STAQLIGSSN	PQHQAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
	301	LIAIGLFLIY	QNGLTLLFEA	VEDGKIHFGL	GLLPMHIIMF	VIAIVLLRVR
	351	SMPSQPFWQA	VG...			

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

50	1	ATGATTATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CCGCCGTCGG
	51	CATTTTCGTC	GTCCTCTTGG	CGGTGTTGGT	GTCCACGCAG	GCGATCAACC
55	101	TGCTTGCCG	CGCAGCTGAC	GGGCGTGTCG	CCATCGATGC	CGTGTGGCC
	151	TTAGTCGGCT	TCTGGGTCAT	CGGTATGACC	CCGCTTTTGC	TGGTGTGAC
	201	CGCATTATC	AGCAGCTGA	CCGTATTGAC	CCGCTACTGG	CGCGACAGCG
	251	AAATGTCGGT	CTGGCTATCC	TGCGGATTGG	CGTTGAAACA	GTGGATACGC
60	301	CCCGTCATGC	AGTTTGCCGT	GCCGTTTGCC	ATCCTGATTG	CCGTCATGCA
	351	GCTTTGGGTG	ATACCGTGCG	CAGAGCTGCG	CAGCCGCGAA	TATGCCGAAA
	401	TTTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGGAAGCCGG	CGAGTTCAAT
	451	AACTTGGGCA	AGCGCAACGG	CAGggtttaT	TtcgtcgaaA	CCTTTGACAC
	501	CGaatccgGC	ATCATGA AAA	ACCTGTtctt	GcGCGAACAG	GACAAAAACG
	551	gcggcgacaA	CATCATCTTC	GCcaaaGAag	gtaactTctc	gctgaaggac

-298-

5 601 AACAAAcgca cgctcgaATT GCGCCACGGC TACCGTTACA GCGGcaccgcC  
 651 CGGacGCGCc gactaCAATC AGGTTtcctt cCAAAAacTc aacctgATta  
 701 TCAGCACCAC GCCCAAacTT ATCGaccCCG TTTCCACCG CCGCACCATT  
 751 tcgacCGCCC AAcTGATTGG CAGCAGCAAT CCGCAACATC AGGCAGAATT  
 801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTGCTC TGCCTACTCG  
 851 CCGTGCCGCT TTCCTATTTC AACCCGCGCA GCGGACATAC CTACAATATC  
 901 TTGATTGCCA TCGGTTTGT TTTAATTAC CAAAACGGGC TGACCCGTGCT  
 951 TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGAAGTCTGC  
 10 1001 CTATGCACAT CATCATGTTC GTCATCGCAA TCGTACTTCT GCGCGTCCGC  
 1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT  
 1101 GAAAGgcgGA AAATGA

This corresponds to the amino acid sequence <SEQ ID 506; ORF101ng-1>:

15 1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA  
 51 LVGFWVIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR  
 101 PVMQFAVPFA ILIAVMQLWV IPWAEALSRE YAEILKQKQE LSLVEAGEFN  
 151 NLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNIIF AKEGNFSLKD  
 201 NKRTLELRHG YRYSGTPEGRA DYNQVSFQKL NLIISTTPKL IDPVSHRRTI  
 251 STAQLIGSSN PQHQAELMWR ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI  
 301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFVL GLLPMHIIMF VIAIVLLRVR  
 351 SMPSQFFWQA VGKSLTLKGG K\*

ORF101ng-1 and ORF101-1 show 97.6% identity in 371 aa overlap:

		10	20	30	40	50	60
25	orf101-1.pep	MIYQRNLIKEL	SFTAVGIFV	VLLAVLVSTQ	AINLLGRAAD	GRVAIDAVLA	LVGFWVIGMT
	orf101ng-1	MIYQRNLIKEL	SFTAVGIFV	VLLAVLVSTQ	AINLLGRAAD	GRVAIDAVLA	LVGFWVIGMT
		10	20	30	40	50	60
30	orf101-1.pep	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWLS	SCGLALKQWIR	PVMQFAVPFA	VLVAVMQLWV
	orf101ng-1	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWLS	SCGLALKQWIR	PVMQFAVPFA	ILIAVMQLWV
		70	80	90	100	110	120
35	orf101-1.pep	IPWAEALSRE	YAEILKQKQE	LSLVEAGEFN	SLGKRNGRVY	FVETFDTESG	IMKNLFLREQ
	orf101ng-1	IPWAEALSRE	YAEILKQKQE	LSLVEAGEFN	SLGKRNGRVY	FVETFDTESG	IMKNLFLREQ
		130	140	150	160	170	180
40	orf101-1.pep	DKNGGDNIIF	AKEGNFSLND	NKRTLELRHG	YRYSGTPEGRA	DYNQVSFQKL	NLIISTTPKL
	orf101ng-1	DKNGGDNIIF	AKEGNFSLND	NKRTLELRHG	YRYSGTPEGRA	DYNQVSFQKL	NLIISTTPKL
		190	200	210	220	230	240
45	orf101-1.pep	IDPVSHRRTI	PTAQLIGSSN	PQHQAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
	orf101ng-1	IDPVSHRRTI	STAQLIGSSN	PQHQAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
50		250	260	270	280	290	300
55	orf101-1.pep	LIAIGLFLIY	QNGLTLLFEA	VEDGKIHFVL	GLLPMHIIMF	FAVALILLRVR	SMPSQFFWQA
	orf101ng-1	LIAIGLFLIY	QNGLTLLFEA	VEDGKIHFVL	GLLPMHIIMF	VIAIVLLRVR	SMPSQFFWQA
		310	320	330	340	350	360
60	orf101-1.pep	VGKSLTLKGG	KX				
	orf101ng-1	VGKSLTLKGG	KX				
		370					

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

```

5      1  ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
      51  GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
     101  TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACCT
     151  ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT
     201  TCGTGTCTGC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
    10  251  CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC
     301  AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
     351  TAT.GTATGC CAACAAAATC ACCTTGATCA GTACGGTCTGA GCAAGCAGGC
     401  ATTCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```

15      1  ..GGGFINASCA TLTTAKPQYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
      51  ILSYHSKIDA PVWQDVRVV AGQNDVAATG DAHSPILNNA AANTSNTAN
     101  NGTHIPLFAI DTGKLGXVC QQNHLQYGR ASRHS*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with with *pspA* putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and *pspA* show 44% aa identity in 179aa overlap:

```

      orf113  GGGFINASCATLTTAKPQYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
      GGG INA+ TLT+ P G+L+ F + G VVI G GLD D DYTRILS ++I+A
      pspA    GGGLINAASVTLTSGVPLNNGNLGTGFDVSSGKVVIGKGLDTSADYTRILSRAAEINA 256

    25      orf113  PVWQDVRVVAGQNDVAATGDAHSPILXXXXXXXXXXXXXXXXXGTHIPLFAIDTGLKLGMYA 120
      VWG+DV+VV+G+N + G + P AIDT LGGMYA
      pspA    GVWQKDVKVVS GKNKLD FDG-----SLAKTASAPSSSDSVTPTVAIDTATLGGMYA 307

    30      orf113  NKITLISTVEQAGIRNQGFASAGNVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
      +KITLIST A IRN+G+ FA+ G V ++A+GKL N+G I A +++ A+ V N
      pspA    DKITLISTDNGAVIRNKRIFAATGGVTLSDAGKLSN GSGSIDAA----EITISAQTVDN 362

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

```

      orf113                                     GGGFINASCATLTTAKPQYQAGDLSAFKIR 30
      orf113ng  SHPSQLNGYIEVGRRRAEVVIANPAGIAVNGGGFINASRATLTTCGPQYQAGDFSGFKIR 224
      orf113    ||||| |||||:|||||:|:|
      orf113ng  QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWQDVRVVAGQNDVAATGDAHSPILNNA 90
      orf113    |||:|||||:|||||:|
      orf113ng  QGNAVIAGHGLDARDTDFTRILVCQQNHLQYGRTSRHS 263
      orf113    IDTGKLGXVCQQNHLQYGRASRHS 135
      orf113ng  DFSGFKIRQGNVVIAGHGLDARDTDFTRILVCQQNHLQYGRTSRHS 263

```

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

```

50      1  MNKTLRYVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIFTH
      51  SKAFCFSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNIGP

```

-300-

```

101 QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
151 TRGEARVVVN QINSSHPSQL NGYIEVGRR AEVVIANPAG IAVNGGGFIN
201 ASRATLTGQ PQYQAGDFSG FKIRQGNVAGHGLDARDT DFTRILVCQQ
251 NHLDQYGRS RHS*

```

- 5 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 61

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

```

10      1  .TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
      51  CAACATTTCA CTGGGTTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTAA
     101  GCCATCATGC GCCCAGCCAA GGCACGTAGT TGCCGCAAAG CAACGGTATT
     151  TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
     201  ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTTGAA ACCGATCCAC
     251  GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGACAGC
     15  301  CTCAAAC TAG ACCCAAACA TTTACATAAA CGTTTGGGTG ATGTTATTA
     351  CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
     401  GTTTAGAcGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
     451  AATGGCGCGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGCATT
     501  AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
     20  551  AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTTGGTGCCA
     601  CAGGTTTATG TACCGGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT
     651  GTTGT CAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
     701  CAGGCACGAT TGCAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA
     751  GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
     25  801  ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT
     851  TGCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT
     901  CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
     951  TATCACAGGC AAAGAAAAAG GTGTTT..

```

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

```

30      1  .STGHSEQNYT LPREITRNIS LGSFAYESHR KALSHHAPSQ GTELPQSNIGI
     51  SLPYTSNSFT PLPSSSLYII NPVNKGYLVE TDPRFANYRQ WLGS DYMLDS
     101  LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
     151  NGATAARSMN LSVGIALSAE QVAQLTSDIV WLQKEVKLP DGGTQTVLVP
     201  QVYVRVKNKD IDGKGALLSG SNTQINVS GS LKNSGTIAGR NALIINTDTL
     35  251  DNIGGRIHAQ KSAVTATQDI NNIGGMLSAE QTLNLAGNN INSQSTTASS
     301  QNTQGSSTYL DRMAGIYITG KEKGV..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and *pspA* protein show 50% aa identity in 325aa overlap:

```

40      Orf115: 1  STGHSEQNYTLPREITRNISLGSFAYESHKALSHHAPSQGTLPQSNIGISLPYTSNSFT 60
           STG+S  Y  E++ +I +G AY+ +  +  P  +  NGI  +T
      pspA:  778  STGYSRSPYEPAPEVS-SIRMGISAYKGYAPQQASDIPGTVPVVAENGIHPTFT----- 831

      Orf115: 61  PLPSSSLYIINPVNKGYLVE TDPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQR 120
           LP+SSL+ I P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+
     45  pspA:  832  -LPNSSLFAIAPNNKGYLIETDPAFTDYRKWLGSGYMLAALQQDPNHHKRLGDGYEQK 890

      Orf115: 121  LINEQIAELTGHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180
           L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV
     50  pspA:  891  LVNEQIAKLTYRRLDGYTNDEEQFKALMDNGITIAKELQLTGIALSAEQVARLTSDIV 950

      Orf115: 181  WLQKEVKLPDGGTQTVLVPQVYVRVKNKDIDGKGALLSGSNTQINVS GS LKN-SGTIAG 239
           WL  + V LPDG TQTVL P+VYVR +  D++G+GALLSGS I SG+++N G IAG
     55  pspA:  951  WLENETVTLPDGTTQTVLKP KYVVRARPKDMNGQALLSGSVVDIG-SGA IENRGGLIAG 1009

```



Orf115: 240 RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGGMLSAEQTLNAGXXXXXXXXXXXX 299  
R ALI+N + N+ G + + A DI N G + AE LLL A  
pspA: 1010 REALILNAQNIKNLQGD LQGNIFAAAGSDITNTGS-IGAENALLLKASNNIESRSETRS 1068

5 Orf115: 300 XXXXXXXXXXXYLD RMAGIYITGKEKG 324  
+ R+AGIY+TG++ G  
pspA: 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093

Homology with a predicted ORF from *N.gonorrhoeae*

10 ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from  
*N.gonorrhoeae*:

15	orf115.pep	STGHSEQNYTLPREITRNLISLGSFAYESHRK	31
	orf115ng	NEQTFGKKVFSENGLKLNHYWRARRKGHDETGHREQNYTLPPEITRDISLGSFAYESHK	71
20	orf115.pep	ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYLVET	81
	orf115ng	ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSSLYIINPANKGYLVET	131
25	orf115.pep	DPRFANYRQWLGS DYMLDSLKLPNNLHKRLG DGYEQRLINEQIAELTGHRRLDGYQND	141
	orf115ng	DPRFANYRQWLGS DYMLGSLKLPNNLHKRLG DGYEQRLINEQIAELTGHRRLDGYQND	191
30	orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWL VQKEVKLPDGGTQT VLVLPQ	201
	orf115ng	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLPDGGTQT VLVLPQ	251
35	orf115.pep	VYVRVKNGDIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	261
	orf115ng	VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	311
40	orf115.pep	SAVTATQDINNIGGMLSAEQTL LLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK	321
	orf115ng	SAVTATQDINNIGGILSAEQTL LLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK	371
	orf115.pep	EKGV	325
	orf115ng	EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDNLDTVQTGKYQEIHFDADNHTIR	431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino acid sequence <SEQ ID 514>:

	1	MLVQTEKDGL	HNEQTFGEEK	VFSENGKLHN	YWRARRKGHD	ETGHREQNYT
	51	LPEEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDRPRFANYRQ	WLGSDYMLGS
45	151	LKLDPNNLHK	RLGDGYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLVOKEVKLP	DGGTQTVLMP
	251	QVYVRVKNGG	IDGKGALLSG	SNTQINVSGS	LKNSGTTIAGR	NALININTDTL
	301	DNIGGRIHAQ	KSAVATATQDI	NNIGGILSAE	QTLLLNAGNN	INNQSSTAKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQAQ	KGDINIAGQ	ISNQSQGGQT
50	401	RLQAGRDINL	DTVQTGKYQE	IHFADADNHTI	RGSTNEVGSS	IQTKGDVTL
	451	SGNNLNAKAA	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDS	ASKHTGRSGG
	501	GKNKLVTDKA	QSHHETAQSS	TFEKGQVVLT	AGNADANGLD	NVISDNGTRI
	551	QAGNHVRIGT	TQTQSQSEY	HQTQKSGLMS	AGIGFTIGSK	TNTQENQSQS
	601	NEHTGSTVGS	LYQDTTIVAS	KHYEQTGSNV	SSPEGNNLIS	TQSMOIGAAQ
55	651	NQLNSKTTQT	KEQDKGLTVAF	SSPVTDLAQ	AIAVAHA	QFDKAKTTAL
	701	MPWRLPMOVG	RLFKQAKAPK	K*		

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

60

1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
51	CGAGAGAAAA	GTCTTCAGCG	AAAATGGTAA	GTTCGACAAC	TACTGGCGGT
101	CGCGTCGTAA	AGGACATGAT	GAACAGGGC	ATCGTGACA	AAATTACT
151	TTGCCGGAGG	AAATCACACG	CGACATTTCA	CTGGGTTCAT	TGCGCTATGA
201	ATCGCATAGC	AAAGCATTAA	GCCGTCATGC	GCCGAGCCAA	GGCATTGAGT
251	TGCCACAAAG	TAACCGGGAT	AATATCCGTA	CTCGGAAAAG	CAACGGTATT

-302-

301 TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT  
 351 ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC  
 401 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC  
 451 CTCAAACCTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA  
 501 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC  
 551 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT  
 601 AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATGTGCATT  
 651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC  
 701 AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA  
 751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT  
 801 GTTGTGAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT  
 851 CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA  
 901 GACAATATCG GTGGCGGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC  
 951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT  
 1001 AATCTCAATG GGGTAACAAC ATCAACAACC AAAGCAGCGC CAAGAGCAGT  
 1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA  
 1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA  
 1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC  
 1201 CGGTGCGAGC CAGGACGCGA CATTAACTTG GATACGGTAC AAACCGGCAA  
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA  
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtatTG  
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAAGTCGCA GCGCAAAAGG  
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC  
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC  
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC  
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT  
 1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAACTC AAAGCCAAAG  
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG  
 1751 GCTTCACTAT TGGCAGCAAG ACAACACAC AAGAAAACCA ATCCCAAAGC  
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT  
 1851 TGTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG  
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA  
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC  
 2001 GGTGGCATTC AGTTGCCCCG TTACCGATTT GGCACAACAA GCGATTGCCG  
 2051 TAGCACACAA AGCAGCAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA  
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA  
 2151 GGCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

40 1 LLVQTEKDGL HNEQTFGEKK VFSNGKLHN YWRARRKGHD ETGHREQNYT  
 51 LPEEITRDIS LGSFAYESHK KALSRHAPSQ GTELPQSNRD NIRTAKSNGI  
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPFRFANYRQ WLGS DYMLGS  
 151 LKLDPNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP  
 45 251 QVYVRVKNCG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGNAGN INNQSTAKSS  
 351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINI IAGQ ISNQSDQGGT  
 401 RLQAGRDLNL DTVQTKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL  
 451 SGNLNLAKAA EVGSAKGTLL VYAKNDITIS SGIHAGQVDD ASKHTGRSGG  
 50 501 GNKLVIDDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVIDSNGTRI  
 551 QAGNHVRIGT TQTQSQSEY HQTKSGLMS AGIGFTIGSK TNTQENQSQS  
 601 NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNNLIS TQSM DIGAAQ  
 651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL  
 701 MPWRLPMQVG RPIKQAKAHK T\*

55 This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

20 30 40 50 60 70  
 orf115ng-1.p NEQTFGEKKVFSNGKLHNYWRARRKGHDSTGHREQNYTLPEEITRDISLGSFAYESHK  
 orf115 STGHSEQNYTLPREITRNISLGSFAYESHK  
 10 20 30  
 80 90 100 110 120 130  
 orf115ng-1.p ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET  
 orf115 ALSRHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGIVET  
 40 50 60 70 80

-303-

5 orf115ng-1.p 140 150 160 170 180 190  
 orf115 140 150 160 170 180 190  
 90 100 110 120 130 140

10 orf115ng-1.p 200 210 220 230 240 250  
 orf115 200 210 220 230 240 250  
 150 160 170 180 190 200

15 orf115ng-1.p 260 270 280 290 300 310  
 orf115 260 270 280 290 300 310  
 210 220 230 240 250 260

20 orf115ng-1.p 320 330 340 350 360 370  
 orf115 320 330 340 350 360 370  
 270 280 290 300 310 320

25 orf115ng-1.p 380 390 400 410 420 430  
 orf115 380 390 400 410 420 430  
 EKVGLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFADADNHTIR  
 EKV

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

30 gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length  
 = 2273  
 Score = 604 bits (1541), Expect = e-172  
 Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

35 Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHDETGHRQNYTLPEEITRDIS 60  
 L+V T + L N++T G K + ++ G L H Y R +KG D TG+ Y E++ I  
 Sbjet: 739 LIVGTPESALDNDLGTGTI-TDKGDLHRYHRHKKGRDSTGYSRSPYEPAPEVS-SIR 796

40 Query: 61 LGSFAYESHKALSRRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120  
 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I  
 Sbjet: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGHPTFT-----LPNSSLFAI 840

45 Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELT 180  
 P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+L+NEQIA+LT  
 Sbjet: 841 APNNKGYLIETDPAFTDYRKWLGSYMLAALQQDPNHIHKRLGDGYEQKLVNEQIAKLT 900

50 Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLP 240  
 G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP  
 Sbjet: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQLTGIALSAEQVARLTSDIVWLENETVTLF 960

55 Query: 241 DGGTQTVLMPQVYVRVKNGGIDGKALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299  
 DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N  
 Sbjet: 961 DGTQTVLKPVKVYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

60 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTY 359  
 + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS  
 Sbjet: 1020 IKNLQGDLDQGNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078

65 Query: 360 LDRMAGIYITGKEKVGLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQ 419  
 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q  
 Sbjet: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138

70 Query: 420 EIHFADADNHTIRGSTNEVGSSIQTKGDVTLNAGNNLNKAAAEVGSAGKTLAVYAKNDITI 479  
 FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +  
 Sbjet: 1139 NTIFSDNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQRLKLAAGRDIV 1198

Query: 480 SSGIHAGQVDDASKHTGRSGGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539  
 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G  
 Sbjet: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMTIRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258

-304-

Query: 540 SNVISDNGTRIQAAGNHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598  
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S  
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318

5 Query: 599 QSNEHTGSTVGSGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658  
 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++  
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISGKISIDAAQNRYSQESK 1378

10 Query: 659 QTYEQKGLTVAFSSPVT D 676  
 Q YEQKG+TVA S PV +  
 Sbjct: 1379 QVYEQKGVTV AISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 62

15 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

1 ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG  
 51 TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA  
 101 ACACGACCCA TGTGTATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT  
 151 GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC  
 20 CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 251 ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT  
 301 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAACCTC AAAGCCAAAG  
 351 CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG  
 401 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC  
 25 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT  
 501 TGTTGCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG  
 551 AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC  
 601 AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAGG.CTAAC  
 651 GGTGGCATT C AGTTCCGCCG TTACCGATT GGCACAACAA ...

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

1 ..SGNNLNAKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG  
 51 GNKLVIDTKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTQI  
 101 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS  
 151 NEHTGSTVGS LKGDTTIVAG KHYEQIGSTV SPEGNNNTIY AQSIDIQAAH  
 35 201 NKLNSNTTQT YEQKXLTVAF SSPVTDLAQQ ...

Computer analysis of this amino acid sequence gave the following results:

Homology with the *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and *pspA* protein show 45% aa identity in 224aa overlap:

Orf117: 4 NLNAKAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGKNKLVIDTKAQSH 63  
 ++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++  
 pspA: 1173 DIRIRAAEVGSEQGRKLKLAAGRDIKVEAGKAHTETEDALKYTGRSGGGIKQKMT RHLKNQ 1232

Orf117: 64 HETAQSSTFEGKQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSEYHQT 123  
 + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++  
 45 pspA: 1233 NGQAVSGTLDGKEIILVSGRDI TVTGSNIIADNHTILSAKNNIVLKAETRSRSAEMNKK 1292

Orf117: 124 QKSGLM-SAGIGFTIGSKTNTQENQSQSNEHTGSTVGSGLKGDTTIVAGKHYEQIGSTVSS 182  
 +KSGLM S GIGFT GSK +TQ N+S++ HT S VGS L G+T I AGKHY Q GST+SS  
 pspA: 1293 EKSGLMGSGGIGFTAGSKKDTQTNRSETVSHTESVVGSLNGNTLISAGKHYTQTGSTISS 1352

Orf117: 183 PEGNNNTIYAQSIDIQAAHNKLNSNTTQTYEQKXLTVAFSSPVT D 226  
 P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +  
 50 pspA: 1353 PQGDVGISGKISIDAAQNRYSQESKQVYEQKGVTV AISVPVVN 1396

Homology with a predicted ORF from *N.gonorrhoeae*

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from *N.gonorrhoeae*:

5	orf117.pep	SGNNLNAKAAEVSSANGTLAVSANNDINIS	30
	orf117ng	IHFADNHTIRGSTNEVGSSIQTGKDVTLSSGNNLNAKAAEVGSAKGT LAVYAKNDITIS	480
10	orf117.pep	AGINTTHVDDASKHTGRSGGKNLVIDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	90
	orf117ng	SGIHAGQVDDASKHTGRSGGKNLVIDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	540
15	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSETYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	150
	orf117ng	NVISDNGTRIAGNHVRIGTTQTQSQSETYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	600
20	orf117.pep	NEHTGSTVGS LKGD TTIVAGKH YE QIGSTVSSPEGNN TIYAQSIDIQA AHNKLSNTTQT	210
	orf117ng	NEHTGSTVGS LKGD TTIVASKH YE QTSNVS SPEGNN LI STQSM DIGAAQNQLNSKTTQT	660
25	orf117.pep	YEQKXLTVA FSSPVTDLAQQ	230
	orf117ng	YEQKGLTVAFSSPVTDLAQQA IAVAHKAAKQFDKAKTTALMPWRLPMQVGR LFKQAKAPK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

25	1	..LLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHREQNYT
	51	LPEEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
30	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDFRFANYRQ	WLGSDYMLGS
	151	LKLDPNNLHK	RLGDGYYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
35	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLVQKEVKLP	DGGTQTVLMP
	251	QVYVRVKNGG	IDGKGALLSG	SNTQINVSGS	LKNSGTIAGR	NALIINTDTL
40	301	DNIGGRIHAQ	KSAVTATQDI	NNIGGILSAE	QTLNLLNAGN	INNQSTAKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIIAGQ	ISNQSDQGGT
45	401	RLQAGRDINL	DTVQTGKYQE	IHFADNHTI	RGSTNEVGSS	IQTGKDVTL
	451	SGNNLNAKAA	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
50	501	GNKLVIDDKA	QSHHETAQSS	TFEGKQVVLQ	AGNDANILGS	NVISDNGTRI
	551	QAGNHVRIGT	TQTQSQSETY	HQTQKSGLMS	AGIGFTIGSK	TNTQENQSQS
55	601	NEHTGSTVGS	LKGD TTIVAS	KHYEQTGSNV	SSPEGNN LIS	TQSM DIGAAQ
	651	NQLNSKTTQT	YEQKGLTVAF	SSPVTDLAQQ	AI AVAHKAAK	QFDKAKTTAL
	701	MPWRLPMQVG	RLFKQAKAPK	K*		

40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

45	1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
	51	CGAGAAGAAA	GTCTTCAGCG	AAAATGGTAA	GTTGCACAAC	TACTGGCGTG
50	101	CGCGTCGTAA	AGGACATGAT	GAAACAGGGC	ATCGTGAACA	AAAT TATACT
	151	TTGCCCGAGG	AAATCACACG	CGACATTTCA	CTGGGTTTCAT	TTGCCATATGA
55	201	ATCGCATAGC	AAAGCATTA	GCCGTCATGC	GCCCAGCCAA	GGCACTGAGT
	251	TGCCACAAAG	TAACCGGGAT	AATATCCGTA	CTGCGAAAAG	CAACGGTATT
60	301	TCGCTACCCT	ATACGCCCAA	TTCTTTTACC	CCATTACCCG	GCAGCAGCTT
	351	ATACATTATC	AATCTGCGCA	ATAAAGGCTA	TCTTGTTGAA	ACCGATCCAC
65	401	GCTTTGCCAA	CTACCGTCAA	TGGTTGGGTA	GTGACTATAT	GCTGGCGAGC
	451	CTCAAACTAG	ACCCAAACAA	TTTACATAAA	CGTTTGGGTG	ATGGTTATTA
70	501	CGAGCAACGT	TTAATCAATG	AACAAATCGC	AGAGCTGACA	GGGCATCGTC
	551	GTTTAGACGG	TTATCAAAAC	GACGAAGAAC	AATTTAAAGC	CTTAATGGAT
75	601	ATGGCGCGA	CTGCGGCACG	TTCGATGAAT	CTCAGCGTTG	GCATGTCATT
	651	AAGTGCCGAG	CAAGCAGCGC	AACTGACCAG	CGATATTGTT	TGGTTGGTAC
80	701	AAAAAGAAGT	TAAACTTCCT	GATGGCGGCA	CACAAACCGT	ATTGATGCCA
	751	CAGGTTTATG	TACGCGTTAA	AAATGGCGGC	ATAGACGGTA	AAGGTGCATT
85	801	GTTGTCAAGC	AGCAATACAC	AAATCAATGT	TTCAGGCAGC	CTGAAAAACT
	851	CAGGCACGAT	TGCAGGGCGC	AATGCGCTTA	TTATCAATAC	CGATACGCTA
90	901	GACAATATCG	GTGGGCGTAT	TCATGCGCAA	AAATCAGCGG	TTACGGCCAC
	951	ACAAGACATC	AATAATATTG	GCGGCATTCT	TTCTGCCGAA	CAGACATTAT
95	1001	TGCTCAATGC	GGGTAACAAC	ATCAACAACC	AAAGCACGGC	CAAGAGCAGT
	1051	CAAAATGCAC	AAGGTAGCAG	CACCTACCTA	GACCGAATGG	CAGGTATTTA

1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA  
 1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC  
 1201 CGGCTGCAGG CAGGACGCGA CATTAACCTG GATACGGTAC AAACCGGCAA  
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCOA  
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCTatTG  
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG  
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC  
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC  
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC  
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATT CCGATAATGG CACCCGATT  
 1651 CAAGCAGGCA ATCATGTTTCG CATTTGGTACA ACCCAAATC AAAGCCAAAG  
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG  
 1751 GCTTCACTAT TGGCAGCAAG ACAACACAC AAGAAAACCA ATCCCAAAGC  
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCCATCC  
 1851 TGTTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG  
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGACAAA  
 1951 AACCATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC  
 2001 GGTGGCATTC AGTTGCCCG TTACCGATT GGCACAACAA GCGATTGCCG  
 2051 TAGCACACAA AGCAGCAAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA  
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA  
 2151 GGCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT  
 51 LPEEITRDIS LGSFAYESHK KALSRHAPSQ GTELPQSNRD NIRTAKSNGI  
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPREFANYRQ WLGSDYMLGS  
 151 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP  
 251 QVYVRVKNNG IDGKGALLSG SNTQINVS GS LKNSGTIAGR NALIINTDTL  
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGN NINNSTAKSS  
 351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINIAGQ ISNQSDQGQT  
 401 RLQAGRDI NL DTVQTKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLT  
 451 SGNNLNAKAA EVGSAGKTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG  
 501 GNKLIVITDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTRI  
 551 QAGNHVRIGT TQTQSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS  
 601 NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMIDGAAQ  
 651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL  
 701 MPWRLPMQVG RPIKQAKAHK T\*

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it  
 shows homology with a secreted *N.meningitidis* protein in the database:

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis]Length = 2273  
 Score = 604 bits (1541), Expect = e-172  
 Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)  
 Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHD ETGHREQNYTLPEEITRDIS 60  
 L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I  
 Sbjct: 739 LIVGTPESALDNDDELGTCTI-TDKGDLHRYHRHHKKGRDSTGYSRSPYEPAPEVS-SIR 796  
 Query: 61 LGSFAYESHKALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120  
 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I  
 Sbjct: 797 MGISAYKGY-----APQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840  
 Query: 121 NPANKGYLVETDPREFANYRQWLGSDYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELT 180  
 P NKGYL+ETDP F +YR+WLG YML +L+ DPN++HKRLGDGYEQL+L+NEQIA+LT  
 Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGSYMLAALQQDPNHIHKRLGDGYEQLVNEQIAKLT 900  
 Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVQKEVKLP 240  
 G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP  
 Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQTPGIALSAEQVARLTSDIVWLENETVTLP 960  
 Query: 241 DGGTQTVLMPQVYVRVKNNGIDGKGALLSGSNTQINVSGLKN-SGTIAGRNALIINTDT 299  
 DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N  
 Sbjct: 961 DGTQTVLKPQVYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019  
 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNINNSTAKSSQNAQGSSTY 359

-307-

+ N+ G + + A D I N G I A E L L L A N N I ++S +S+QN QGS  
 Sbjct: 1020 IKNLQGD LQGNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078  
 Query: 360 LDRMAGIYITGKEKGVLAQAQAGKDINIIAGQISNQSDQGQTR LQAGRDINLDTVQTGKYQ 419  
 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q  
 Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138  
 Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTL LSGNNLNAKAAEVGSAKGT LAVYAKNDITI 479  
 FD+DN+ IR NEVG S+I+T+G+++L + ++ +AAEVGS +G L + A DI +  
 Sbjct: 1139 NTIFDS DNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGR LKLAAGRDIKV 1198  
 Query: 480 SSGIHAGQVDDASKHTGRSGGGNKL VITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539  
 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G  
 Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMTRHLKNQNGQAVSGTLDGKEIILVSGR DITVTG 1258  
 Query: 540 SNVISDNGTRIQA GNVHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598  
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S  
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKA AETR SRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318  
 Query: 599 QSNEHTGSTVGS LKGD TTIVASKHYEQTGSNVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658  
 ++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++  
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRY SQESK 1378  
 Query: 659 QTYEQKGLTVAFSSPVT D 676  
 Q YEQKG+TVA S PV +  
 Sbjct: 1379 QVYEQKGVTV AISVPV N 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 63

30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA  
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG  
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAAC CAG CCATGTCCGC  
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACC GGC  
 35 201 GGTCAAAAA ACGGCAAAC CCCAAGACCC CGYCATGCGC AACCTGCAAG  
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAA AGCCTCCCCG  
 301 TTCAA AACC G AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA  
 351 CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC  
 401 CTGCCGACGC GTCGGCAAAC CCTGCACCCG TTCCGCAAAC ACCTGCAAAA  
 40 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA  
 501 CGTGCGCATC GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSXTSHVR  
 51 DGKPSGG SVM MPKPQPAVKK TAKPQDPXMR NLQE QDAVYI AKQKQAKASP  
 45 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS ATKPADASAK PAPVPQTPAK  
 151 PLITLKELSK VELSWFDVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA  
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG  
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAA AACCAG CCATGTCCGC  
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACC GGC  
 50 201 GGTCAAAAA ACGGCAAAC CCCAAGACCC CGCCATGCGC AACCTGCAAG  
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAA AGCCTCCCCG  
 301 TTCAA AACC G AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA  
 55 351 CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACCGAAAC  
 401 CTGCCGACGC GCGGCAA AAA CCTGCACCCG TTCCGCAAAC ACCTGCAAAA  
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAC CCTGGTTTGA  
 501 CGTGCGCTTC GACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC  
 551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTCCGC  
 60 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG

5 651 CTATCAGGCA TTTATCGTGG GTATTCAGGC AGTCAGCCGC AACGGACTTG  
 701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA CGCATTCGCA  
 751 CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CTTTATCGA  
 801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG  
 851 CCATCCATTT GGTTTCCCGG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC  
 901 GCCGTAAACGG GCGTGGGTTT CGTTTGGAA GACGACGGCG CGTTCCTACTA  
 951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG  
 1001 AGCCGTTTAC CAACGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT  
 1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAA CCTTCGACGA  
 1101 TTTGTTTATG GATTTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC  
 1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG  
 1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA  
 1251 ACCGGGCGGC AAAACCGCAT TGGCCTGTT CTCCTAA

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

15 1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR  
 51 DGKPSGGSV MPKPQPAVKK TAKPQDPAMR NLQEQDAVYI AKQKQAKASP  
 101 FKTEIETALE ESGIIGNSAH TVSEFQTGHS APKPADAPAK PAPVPQTPAK  
 151 PLITLKELSK VELPWFVDFR DFISYIALTE AKELHALPRL SNRCRYQIVG  
 201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA  
 251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS  
 301 AVTGVGVFLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS  
 351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV  
 401 RTYVRLARQSE MLKVGIEPGG KTAIRLFS\*

Computer analysis of this amino acid sequence gave the following results:

#### 25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
30	orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRD					
	orf119a	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRD					
		10	20	30	40	50	60
35	orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
	orf119a	MPKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
		70	80	90	100	110	120
40	orf119.pep	TVSEFQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFVDFRIDFISY					
	orf119a	TVPEPQTGHSAPKPADAPAKPVFPQTPAKPLITLKELSKVELPWFVDFRDFISYIALTE					
		130	140	150	160	170	180
45	orf119a	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS					
		190	200	210	220	230	240

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

50 1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTGCGCTA  
 51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG  
 101 GGCACCTCCGA CAAAGATGCC CTGCTCAACA GCAAACCAG CCATGTCGCG  
 151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCAACCGGC  
 201 GGTCAAAAAA ACGGCAAAAT CCAAGACCC CGCCATGCGC AACCTGCAAG  
 251 AGCAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCG  
 55 301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGC  
 351 CTCGCCCCAC ACCGTTCCCG AACCCTAAAC CGGACATTCC GCACCAAAAC  
 401 CTGCCGACGC GCCGGCAAAA CCTGTTCCCG TTCCGCAAAAC GCCGGCAAAA  
 451 CCGCTGATTA CGCTCAAAGA GCTGTCGAAG GTCGAGCTGC CCTGGTTTGA  
 501 CGTGCGCTTC GACTTCATCT CTTATATCGC GCTGACCGAA GCCAAAGAAC  
 60 551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTCGCG  
 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG



-309-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

```

651 CTATCAGGCA TTTATCGTGG GTATTCAGGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA TGCATTTCGA
751 CACAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACTATCG
851 CCATCCATT TGGTTTCCCGG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTTGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTATAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TGCGCCTGTT CTCCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 528>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

```

1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
51 DGKPSGGPVM MPKPQPAVKK TAKSQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVPEPQTGHS APKPADAPAK PVPVPQTPAK
151 PLITLKELSK VELPWFDFRF DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 HSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
301 AVTGVGVFVLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQS YKGF S
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVLARQSE MLKVGIEPGG KTALRLFS*

```

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60  
65

```

orfl119a.pep      10      20      30      40      50      60
MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRDGKPSGGPVM
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orfl119-1          10      20      30      40      50      60
MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRDGKPSGGPVM

orfl119a.pep      70      80      90     100     110     120
MPKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orfl119-1          70      80      90     100     110     120
MPKPQPAVKKTAKPQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH

orfl119a.pep     130     140     150     160     170     180
TVPEPQTGHSAPKPADAPAKPVPVQTPAKPLITLKELSKVELPWFDFRFDFISYIALTE
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orfl119-1         130     140     150     160     170     180
TVSEPQTGHSAPKPADAPAKPVPVQTPAKPLITLKELSKVELPWFDFRFDFISYIALTE

orfl119a.pep     190     200     210     220     230     240
AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orfl119-1         190     200     210     220     230     240
AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS

orfl119a.pep     250     260     270     280     290     300
AFNRQVDAFAHSMGGQTLHTDLAAFIEVASALDAFCARVDQTIAIHLVSPSTISGVELRS
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orfl119-1         250     260     270     280     290     300
AFNRQVDAFAQSMGGQTLHTDLAAFIEVASALDAFCARVDQTIAIHLVSPSTISGVELRS

orfl119a.pep     310     320     330     340     350     360
AVTGVGVFVLEDDGAFHYTDTSGSTMFSICS LNNEPFTNALLDNQS YKGF SMLLDIPHSPA
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orfl119-1         310     320     330     340     350     360
AVTGVGVFVLEDDGAFHYTDTSGSTMFSICS LNNEPFTNALLDNQS YKGF SMLLDIPHSPA

orfl119a.pep     370     380     390     400     410     420
GEKTFDDLFM DLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPGG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orfl119-1         370     380     390     400     410     420
GEKTFDDLFM DLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPGG

```

```

orf119a.pep    KTALRLFSX
               |||||
orf119-1       KTALRLFSX

```

### 5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N.gonorrhoeae*:

```

10 orf119.pep    MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRDGPSPGGSV 60
    |||||:|||||
    orf119ng     MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVDRGKPSGGPVM 60

    orf119.pep    MPKPQPAVKKTAKPDQPMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH 120
    |||||:|||||
    orf119ng     MPKPQPAVKKPAKPDQDSAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH 120

15 orf119.pep    TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFVDRIDFISY 175
    |||||:|||||
    orf119ng     TVSEPQTGHSAPKPADAPAKPVFPQTPAKPLITLKELSKVELPWVDFVRFDFISYIALTE 180

```

The complete length ORF119ng nucleotide sequence <SEQ ID 529> is:

```

20      1  ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTCGCCTA
    51  CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGC GC GACCAGTTTCG
    101  GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCGCG
    151  GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCCAACC GGC
    201  GGTCAAAAAA CCGGCCAAAC CCCAAGACTC CGCCATGCGC AACCTGCAAG
25  251  AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
    301  TTCAAAACCG AAATCGAAAC CGCCTTGGA GAAATCGGCA TTATCGGCAA
    351  CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACCGAAAC
    401  CTGCCGACGC GCCGGCAAAC CCGGTTCCCG TTCCGCAAAC GCCGGCAAAC
    451  CCGCTGATTA CGCTCAAAGA GCTGTGGAAG GTCGAGCTGC CCTGGTTTGA
30  501  CGTGCGCTtc gACTTCATCT CCTATATCGC GCTGACCGAA GCCAAGAAGAAC
    551  TGCACGCACT GCCGCGCCTT tccAACCGCT GCCGCTACCA GATTGTCCGGC
    601  TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG
    651  CTATCAGGCA TTTATCGTGG GTATCCAGGC AGTCAGCCGC AACGGACTTG
    701  CCTCGCAGGA AGAATCTCC GCATTCAACC GCCAGGCGGA CGCATTCGCA
35  751  CAAAGCATGG GCGGTGAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
    801  AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGGCTCGAC CAGACCATCG
    851  CCATCCATTT GGTTCGCGC ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
    901  GCCGTAACGG GCGTGGGTTT CGTTTTGGAA GACGACGGCG CGTCCACTA
    951  TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
40  1001  AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
    1051  ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAAA CCTTCGACGA
    1101  TTTGTTTATG GATTTGGCGG TACGCCTGTC CGGTCAGTTG AACCTGAATC
    1151  TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTA
45  1201  CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
    1251  ACCGGGCGGC AAAACCGCCC TCGCCTGTT TTCATAA

```

This encodes a protein having amino acid sequence <SEQ ID 530>:

```

50      1  MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
    51  DGKPSGGPVM MPKPQPAVKK PAKPDQSAMR NLQEQDAVYI AKQKQAKASP
    101  FKTEIETALE EIGIIGNSAH TVSEPQTGHS APKPADAPAK PVPVPQTPAK
    151  PLITLKELSK VELPWDFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
    201  CTMDHDFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQADAFa
    251  QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
    301  AVTGVGFVLE DDGAFHYTDT SGSTMFSICS LNNEFFTAL LDNQSYKGFS
    351  MLLDIPHSPA GEKTFDDLEF DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
55  401  RTYVLARQSE MLKVGIEPGG KTALRLFS*

```

ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

```

60 orf119ng      10      20      30      40      50      60
    MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVDRGKPSGGPVM
    |||||:|||||
    orf119-1     MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVDRGKPSGGSV
    10      20      30      40      50      60

```

-311-

		70	80	90	100	110	120
	orf119ng	MPKPQPAVKKPAKPQDSAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH					
5	orf119-1	MPKPQPAVKKTAKPQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
		70	80	90	100	110	120
	orf119ng	TVSEPQTGHSAPKPADAPAKVPVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE					
10	orf119-1	TVSEPQTGHSAPKPADAPAKVPVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE					
		130	140	150	160	170	180
	orf119ng	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS					
15	orf119-1	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS					
		190	200	210	220	230	240
	orf119ng	AFNRQADAFQSMGGQTLHTDLAAFIEVASALDAFCARVDQTTIAIHLVSPSTISGVELRS					
20	orf119-1	AFNRQVDAFAQSMGGQTLHTDLAAFIEVASALDAFCARVDQTTIAIHLVSPSTISGVELRS					
		250	260	270	280	290	300
	orf119ng	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSINNEPFTNALLDNQSYKGFMSMLLDIPHSPA					
25	orf119-1	AVTGVGVFLEDDGAFHYTDTSGSTMFSICSINNEPFTNALLDNQSYKGFMSMLLDIPHSPA					
		310	320	330	340	350	360
	orf119ng	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPGG					
30	orf119-1	GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPGG					
		370	380	390	400	410	420
	orf119ng	KTALRLFSX					
35	orf119-1	KTALRLFSX					
		429					
	orf119ng	KTALRLFSX					
40	orf119-1	KTALRLFSX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 45 Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

	1	..GCGCGGCACG	GCACGGAAGA	TTTCTTCATG	AACAACAGCG	ACAC.ATCAG
	51	GCAGATAGTC	GAAAGCACCA	CCGGTACGAT	GAAGCTGCTG	ATTTCCTCCA
50	101	TCGCCCTGAT	TTTATTGGTA	GTCGGCGGCA	TCGGCGTGAT	GAACATCATG
	151	CTGGTGTCCG	TTACCGAGCG	CACCAAAGAA	ATCGGCATAC	GGATGGCAAT
	201	CGGCGCGCGG	CGCGGCAATA	TTTtGCAGCA	GTTTTTGATT	GAGGCGGTGT
	251	TAATCTGCGT	CATCGGCGGT	TTGGTCGGCG	TGGGTTTGTC	CGCCGCCGTC
	301	AGCCTCGTGT	TCAATCATTT	TGTAACCGAC	TTCCCGATGG	ACATTTCCGC
	351	CATGTCCGTC	ATCGGCGCGG	TCGCCTGTTT	GACCGGAATC	GGCATCGCGT
55	401	TCGGCTTTAT	GCCTGCCAAT	AAAGCAGCCA	AACTCAATCC	GATAGACGCA
	451	TTGGCACAGG	ATTGA			

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

	1	..ARRHGTEDFFM	NNSDXIRQIV	ESTTGTMKLL	ISSIALISLV	VGGIGVMNIM
	51	LVSVTERTKE	IGIRMAIGAR	RGNIXQQFLI	EAVLICVIGG	LVGVGLSAAV
60	101	SLVFNHFVTD	FPMDISAMSV	IGAVACSTGI	GIAFGFMPAN	KAAKLNPIDA
	151	LAQD*				

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

35 Computer analysis of this amino acid sequence gave the following results:

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

50 Homology with a predicted ORF from *N.meningitidis* (strain A)

[illegible]

-313-

```

5  orf134.pep  ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG
   orf134a    ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG
                                     270      280      290      300      310      320

10  orf134.pep  LVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
   orf134a    LVGVGLSAAVSLVFNHFVTDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
                                     330      340      350      360      370      380

15  orf134.pep  LAQDX
   orf134a    LAQDX

```

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

```

20  1  ATGTCGGTGC  AAGCAGTATT  GGCGCACAAA  ATGCGTTCGC  TTCTGACGAT
   51  GCTCGGCATC  ATCATCGGTA  TCGCTTCGGT  TGTCTCCGTC  GTCGCATTGG
  101  GCAACGGTTC  GCAGAAAAAA  ATCCTTGAAG  ACATCAGTTC  GATAGGGACG
  151  AACACCATCA  GCATCTTCCC  AGGGCGCGGC  TTCGGCGACA  GGCAGCAGCG
  201  CAGGATTAAA  ACCCTGACCA  TAGACGACGC  AAAAATCATC  GCCAAACAAA
  251  GCTACGTTGC  TTCCGCCACG  CCCATGACTT  CGAGCGGCGG  CACGCTGACT
  301  TACCGCAATA  CCGACCTGAC  CGCTTCTTTG  TACGGTGTGG  GCGAACAAATA
  351  TTTTCGACGTG  CGCGGGCTGA  AGCTGGAAC  GGGCGGCTG  TTTGACGAAA
  401  ACGATGTGAA  AGAAGACGCG  CAGGTCGTCG  TCATCGACCA  AAATGTCAAA
  451  GACAAACTCT  TTGCGGACTC  GGATCCGTTG  GGTAAACCA  TTTTGTTCAG
  501  GAAACGCCCC  TTGACCGTCA  TCGGCGTGAT  GAAAAAGAC  GAAAACGCTT
  551  TCGCAATTC  CGACGTGCTG  ATGCTTTGGT  CGCCTATAC  GACGGTGATG
  601  CACCAAATCA  CAGGCGAGAG  CCACACCAAC  TCCATCACCG  TCAAAATCAA
  651  AGACAATGCC  AATACCCAGG  TTGCCGAAA  AGGCTGACC  GATCTGCTCA
  701  AAGCGCGGCA  CGGCACGGAA  GATTTCTTCA  TGAACAACAG  CGACAGCATC
  751  AGGCAGATAG  TCGAAAGCAC  CACCGGTACG  ATGAAGCTGC  TGATTTCCTC
  801  CATCGCCCTG  ATTTCAATTG  TAGTCGGCGG  CATCGGCGTG  ATGAACATCA
  851  TGCTGGTGTC  CGTTACCGAG  CGCACCAAAG  AAATCGGCAT  ACGGATGGCA
  901  ATCGGCGCGC  GCGCGGCGAA  TATTTTGCAG  CAGTTTTTGA  TTGAGGCGGT
  951  GTTAATCTGC  GTCATCGGCG  GTTTGGTCGG  CGTGGGTTTG  TCCGCGCCCG
1001  TCAGCCTCGT  GTTCAATCAT  TTTGTAACCG  ACTTCCCGAT  GGACATTTC
1051  GCCATGTCCG  TCATCGGCGC  GGTGCGCTGT  TCGACCGGAA  TCGGCATCGC
1101  GTTCGGCTTT  ATGCCTGCCA  ATAAAGCAGC  CAAACTCAAT  CCGATAGATG
1151  CATTGGCGCA  GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 536>:

```

45  1  MSVQAVLAHK  MRSLLTMLGI  IIGIASVVS  VALNGSQKK  ILEDISSIGT
   51  NTISIFPGRG  FGDRRSGRIK  TLTIDDAKII  AKQSYVASAT  PMTSSGGTLT
  101  YRNTDLTASL  YGVGEQYFDV  RGLKLETGRL  FDENDVKEDA  QVVVIDQNVK
  151  DKLFADSDPL  GKTILFRKRP  LTVIGVMKKD  ENAFGNSDVL  MLWSPYTTVM
  201  HQITGESHTN  SITVKIKDNA  NTQVAEKGLT  DLLKARHGTE  DFFMNSDSI
  251  RQIVESTTGT  MKLLISSIAL  ISLVVGIGV  MNIMLVSVTE  RTKEIGIRMA
  301  IGARRGNILQ  QFLIEAVLIC  VIGGLVGVL  SAAVSLVFNH  FVTDFPMDIS
  351  AMSVIGAVAC  STGIGIAFGF  MPANKAAKLN  PIDALAQD*

```

50 ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:

```

55  orf134a.pep  MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG
   orf134-1    MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG

  orf134a.pep  FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
   orf134-1    FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

  orf134a.pep  RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
   orf134-1    RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

  orf134a.pep  ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE
   orf134-1    ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE

```

-314-

```

5  orf134a.pep  DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134-1    DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134a.pep  IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
   orf134-1    IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
10  orf134a.pep  STGIGIAFGFMPANKAAKLNPIDALAQDX
   orf134-1    STGIGIAFGFMPANKAAKLNPIDALAQDX

```

Homology with a predicted ORF from *N.gonorrhoeae*

15 ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from *N. gonorrhoeae*:

```

20  orf134.pep  ARHGTEFFMNSDXIRQIVESTTGTMKLL 30
   orf134ng    GESHTNSITVKIKDNANTRVAEKGLAELLKARHGTEFFMNSDSIRQIVESTTGTMKLL 264
   orf134.pep  ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG 90
   orf134ng    ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG 324
25  orf134.pep  LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 150
   orf134ng    LVGVGLSAAVSLVFNHFVTDFFPMDISAAVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 384
30  orf134.pep  LAQD 154
   orf134ng    LAQD 388

```

The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

```

35  1  ATGTCGGTGC AAGCAGTATT GGCACACAAA ATGCGTTCGC TTCTGACCAT
   51  GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCGCTGG
   101 GCAACGGTTC GCAGAAAAAA ATCCTCGAAG ACATCAGTTC GATGGGGACG
   151 AACACCATCA GCATCTTCCC CGGGCGCGGC TTCGGCGACA GCGCGACCGG
   201 CAAAATCAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
   251 GCTACGTGTC CTCGCCACG CCCATGACTT CGAGCGCGCG CACGCTGACC
   301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA
   351 TTTCGACGTG CGCGGGCTGA AGCTGGAAAC GGGGCGGCTG TTGATGAGA
   401 ACGATGTGAA AGAAGACGCG CAAGTCGTCG TCATCGACCA AAATGTCAAA
   451 GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTGTTCAG
   501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAAGAC GAAAACGCTT
   551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
   601 CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
   651 AGACAATGCC AATACCCGGG TTGCCGAAAA AGGGCTGGCC GAGCTGCCTA
   701 AAGCACGGCA CGGCACGGAA GACTTCTTTA TGAACAACAG CGACAGCATC
   751 AGGCAGATGG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCCTC
   801 CATCGCCCTG ATTTTCATTG TAGTCGGCGG CATCGGTGTG ATGAACATTA
   851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
   901 ATCGGCGCGC GCGCGGCGAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
   951 GTTAATCTGC ATCATCGGAG GCTTGGTCGG CGTAGGTTTG TCCGCCGCCG
1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ATTTCCCGAT GGACATTTCTG
1051 GCGGCATCCG TTATCGGGGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
1101 GTTCGGCTTT ATGCTGCCA ATAAGGCAGC CAAACTCAAT CCGATAGATG
1151 CATTGGCGCA GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 538>:

```

60  1  MSVQAVLAHK MRSLLTMLGI IIGIASVVSVALGNGSQKK ILEDISSMGT
   51  NTISIFFGRG FGDRRSQKIK TLTIDDAKII AKQSYVASAT PMTSSGGTILT
   101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
   151 DKLFAДСDPL GKTLIFRKRPLTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
   201 HQITGESHTN SITVKIKDNA NTRVAEKGLA ELLKARHGTE DFFMNSDSI
   251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
   301 IGARRGNILQ QFLIEAVLIC IIGLVGVGL SAAVSLVFNH FVTDFPMDIS

```

351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD\*

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

```

5   orf134ng      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSMGNTNISIFPGRG
    orf134-1      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTNISIFPGRG

10  orf134ng      FGDRRSGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
    orf134-1      FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

15  orf134ng      RGLKLETGR LFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
    orf134-1      RGLKLETGR LFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

20  orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGE
    orf134-1      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGE

25  orf134ng      DFFMNNSDSIRQMVESTTGTMKLLISSIALISLVVGIGVMNIMLVSVTERTKEIGIRMA
    orf134-1      DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGIGVMNIMLVSVTERTKEIGIRMA

    orf134ng      IGARRGNILQQFLIEAVLICIGGLVGVLSAAVSLVFNHFVTDFFPMDISAASVIGAVAC
    orf134-1      IGARRGNILQQFLIEAVLICVIGGLVGVLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC

    orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
    orf134-1      STGIGIAFGFMPANKAAKLNPIDALAQDX

```

30 ORF134ng also shows homology to an *E.coli* ABC transporter:

```

sp|P75831|YBJZ ECOLI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ >gi5
(AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length =
648
Score = 297 bits (753), Expect = 6e-80
35  Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

Query: 1  MSVQAVLAHKMRSLLTMLXXXXXXXXXXXXXGNGSQKKILEDISSMGNTNISIFPGRG 60
      M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+
40  Sbjct: 260 MAWRALAANKMRTLLTMLGIIIGIASVVSIVVVGDAAKQMVLA DIRSIGTNTIDVYPGKD 319

Query: 61  FGDRRSGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
      FGD + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
45  Sbjct: 320 FGDDDPQYQQALKYDDLIAIQKQPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

Query: 121  RGLKLETGR LFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTILFRKRPLTVIGVMKK 179
      G+ G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++
50  Sbjct: 380 YGMTFSEGNFTNQEQLNGRAQVVVLDSNTRRQLFPHKADVGEVILVGNMPARVIGVAEE 439

Query: 180  DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT 239
      ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
55  Sbjct: 440 QQSMFGSSKVLRVWLPYSTMSGRVMGQSWLNSITVRVKEGFSAEAEQQLTRLLSLRHGK 499

Query: 240  EDFFMNNSDSIRQMVESTTGTMKXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEIGIRM 299
      +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EIGIRM
60  Sbjct: 500 KDDFTWNMDGVLKTVKTRTLQLFLTLVAVISLVVGIGVMNIMLVSVTERTREIGIRM 559

Query: 300  AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDFFPMDISAASVIGAVA 359
      A+GAR ++LQQFLIE F+ + + S +++ A
Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAFL 619

Query: 360  CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
      CST GI FG++PA AA+L+P+DALA++
Sbjct: 620 CSTVTGILFGWLPARNAARLDPVDALARE 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 65

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

1   ..GGGACGGGAG CGATGCTGCT GCTGTTTTAC GCGGTAACGA T.CTGCCTTT
51  GGGCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
101 TTTCCTTCCT GATTTTGAAA GAACGGATTT CCGTTTACAC GCAGGCGGTG
151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCCG
201 CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGC GCGATGT
251 CCGGCTGGGC GTATTGAAA GTGCGCGAAC TGTCTTTGGC GGGCGAACC
301 GGCTGGCGCG TCCTGTTTTA CCTTTCCTGT ACAGGTGTGG CGATGTCGTC
351 GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCCTTT CCATCGGCAG
401 TTATCTGTC GTGCATCGGC GTGTCCGCGC TGATTGCCCA ACTGTCGATG
15 451 ACGCGCGCCT ACAAAGTCGG CGACAAATTC ACGGTGCGCT CGCTTTCCTA
501 TATGACCGTC GTTTTTCCG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
601 ATTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

20 1   ..GTGAMLLLFY AVTILPLATG VTLSYSSIF LAVFSFLILK ERISVYTQAV
51  LLLGFAGVVL LLNPSFRSGQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP
101 GWRVVFYLSV TGVAMSSVWA TLTGWHTLSF PSAVYLSLCIG VSALIAQLSM
151 TRAYKVGDKF TVASLSYMTV VFSALSAAFF LGEELEFWQEI LGMCIISAV
201 F*

```

- 25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

1   ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGCG
51  GGCGGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAC
101 AATTGCCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTGCGC TCGGGGCTGC CGCGTATTG CGTCGGGACA mCTTCCGCAC
30 201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCTTTGGC CACTGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTT GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCG TTTACACGCA GGCGGTGCTG CTCCTTGCTT
401 TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTCAGGAA
35 451 ACGGCGGCAC TCGCCGGGCT GGCGGGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAAGTGT CTTTGGCGGG CGAACC CGGCGCGTGC
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCGTGCGT TTGGGCGACG
601 CTGACCGGCT GGCACACCTT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
651 CATCGCGGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCTACA
40 701 AAGTCGGCGA CAAATTCACG GTTGCCCTCG TTTCTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCG ATTTTTTCTG GGCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

- 45 This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```

1   MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMLFS
51  TVALGAAAVL RDXFRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLATGV
101 TLSYSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVLN LNPFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
50 201 LTGWHTLSFP SAVYLSLCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

Computer analysis of this amino acid sequence gave the following results:



Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N. meningitidis*:

```

5      orf135.pep      10      20      30
                        GTGAMLLLFYAVTILPLATGVTLSTSSIF
                        |||||
orf135a      STVALGAAVLRRTDTRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSTSSIF
              50      60      70      80      90      100

10     orf135.pep      40      50      60      70      80      90
                        LAVFSFLILKERISVYTQAVLLLGFGVVLNPSFRSGQETAALAGLAGGAMSGWAYLK
                        |||||
orf135a      LAVFSFLILKERISVYTQAVLLLGFGVVLNPSFRSGQETAALAGLAGGAMSGWAYLK
              110     120     130     140     150     160

15     orf135.pep      100     110     120     130     140     150
                        VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSM
                        |||||
orf135a      VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSM
              170     180     190     200     210     220

20     orf135.pep      160     170     180     190     200
                        TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVFX
                        |||||
orf135a      TRAYKVGDKFTVASLSYMTVVFSALSAAFFLAEELEFWQEILGMCIIILSGILSSIRPTAF
              230     240     250     260     270     280

orf135a      KQRLQSLFRQRX
              290     300

```

30 The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGCG
51  GCGGCGCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
101 AATTTGCCCT CCGCAGCGGC GAATTTGGTCT TTTGGCGCAT GCTGTTTTC
35 151 ACCGTTGCGC TCGGGGCTGC CGCGTATTG CGTCGGGACA CCTCCGCAC
201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGCTGCT GTTTACGCG GTAACGCATC TGCCTTTGGC CACGCGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCCTG TTTACACGCA GCGGTGCTG CTCCTTGCTT
401 TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTCAGGAA
451 ACGGCGGCAC TCGCCGGGCT GCGGGCGGCG GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CCGCAACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTATCGGT TTGGGCGACG
601 CTGACCGGCT GGCACACCTT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
45 651 CATCGGCGTG TCCGCGCTGA TTGCCCACT GTCGATGACG CGCGCCTACA
701 AAGTCGGCGA CAAATTCACG GTTGCCTCGC TTTCCTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCGC ATTTTCTCTG GCCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 544>:

```

1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLS
51  TVALGAAAVL RRDTRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLATGV
101 TLYSTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVLL LNPSFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
55 201 LTGWHTLSFP SAVYLSLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL AEELEFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

```

60  orf135a.pep      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
                        |||||
orf135-1      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL

```

-318-

```

5  orf135a.pep  RRDTFRTPHWKNHLNRSMTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE
   orf135-1    RRDXFRTPHWKNHLNRSMTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE

   orf135a.pep  RISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
   orf135-1    RISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG

10  orf135a.pep  WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSMTRAYKVGDKFT
   orf135-1    WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSMTRAYKVGDKFT

15  orf135a.pep  VASLSYMTVVFSALSAAFFLAEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
   orf135-1    VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

20 *N.gonorrhoeae*:

```

   orf135.pep  GTGAMLLLFYAVTXLPLATGVTLSTSSIF  30
   orf135ng    STVTLGAAVLRRTFRTPHWKNHLNRSMTGAMLLLFYAVTHLPLTTGVTLSTSSIF  335

25  orf135.pep  LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK  90
   orf135ng    LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSGQEPALAGLAGGAMSGWAYLK  395

30  orf135.pep  VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSLSCIGVSALIAQLSM  150
   orf135ng    VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM  455

   orf135.pep  TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVF  201
   orf135ng    TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAAF  506

```

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

```

40  1  MPSEKAFRRH  LRTASFQGLH  LHHFHQKV GK  CGIIGFGIHI  FPTLLPAAQG
   51  ILDIQLGLFR  IDFAALAVYR  RTQVDFIHTV  IDGIASDQAF  SEVVQILRRL
  101  NLGHFTDTHL  IAQARRFIAD  FGNIRPMRRG  EAKTFCRCFR  FDGIDIGHGD
  151  FRQCGHINRL  APGKDCRNGK  RDKVFHTRH  YNQVCLEKTN  CSARKIKFRH
  201  QKQAKTHSTS  LAARFTIRES  LSQRPFMDTA  KKDILGSGWM  LVAAACFTVM
  251  NVLIKEASAK  FALGSGELVF  WRMLFSTVTL  GAAAVLRRTD  FRTPHWKNHL
  301  NRSMTGAM    LLLFYAVTHL  PLTTGVTLST  TSSIFLAVFS  FLILKERISV
45  351  YTQAVLLLG  AGVLLNPS  FRSGQEPAL  AGLAGGAMSG  WAYLKVRELS
   401  LAGEPGWRVV  FYLSATGVAM  SSVWATLTGW  HTLSFPSAVY  LSGIGVSALI
   451  AQLSMTRAYK  VGDKFTVASL  SYMTVVFSAL  SAAFFLGEEL  FWQEILGMCI
   501  IISAAF*

```

Further work revealed the following gonococcal sequence <SEQ ID 547>:

```

50  1  ATGGATACCG  CAAAAAAGA  CATTTTAGGA  TCGGGCTGGA  TGCTGGTGGC
   51  GGCGGCCTGC  TTCACCGTTA  TGAACGTATT  GATTAAAGAG  GCATCGGCAA
  101  AATTGCCCCT  CGGCAGCGGC  GAATIGGTCT  TTTGGCGCAT  GCTGTTTCA
  151  ACCGTACGCG  TCGGTGCTGC  CGCGGTATTG  CGGCGCGACA  CCTCCGCAC
55  201  GCCCCATTGG  AAAAACCCT  TAAACCGCAG  TATGGTCGGG  ACGGGGCGCA
   251  TGCTGCTGCT  GTTTTACGCG  GTAACGCATC  TGCCTTTGAC  AACCGCGGTT
   301  ACCCTGAGTT  ACACCTCGTC  GATTTTTttg  GCGGTATTTT  CCTTCCTGAT
   351  TTTGAAAGAA  CGGATTTCCG  TTTACACGCA  GGCGGTGCTG  CTCCTTGTT
   401  TTGCCGCGCT  GGTATTGCTG  CTTAATCCCT  CGTTCGCAG  CGGTCAGGAA
   451  CCGGCGGCAC  TCGCCGGGCT  GCGGGCGGC  GCGATGTCCG  GCTGGGCGTA
60  501  TTTGAAAGTG  CGCGAAGTGT  CTTTGGCGGG  CGAACC CGGC  TGGCGCGTCG
   551  TGTTTACCT  TCCGCAACC  GCGGTGCGCA  TGTCGTGgt  ttggcgacg
   601  Ctgaccggct  ggCACAccT  GTCCTTcca  tcggcagttt  ATCtgtCGGG

```

-319-

5 651 CATCGGCGTG tccgcgCtgA TTGCCCAaCT GtcgatgAcg cGCGcctaca  
 701 aaGTCGGCGA CAAATTCACG GTTGCCCTCGC tttcctaTat gaccgtcGTC  
 751 TTTTCCGCCC TGTCTGCCGC ATTTTCTTg ggcgaagagc tttTctggCA  
 801 GGAAATACTC GGTATGTGCA TCATTatccT CAGCGGCATT TTGAGCAGCA  
 851 TCCGCCCCAT TGCCTTCAAA CAGCGGCTGC AAGCCCTCTT CCGCCAAAGA  
 901 TAA

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

10 1 MDTAKKDILG SGWMLVAAAC FTVMNVLIKE ASAKFALGSG ELVFWRMFLS  
 51 TVTLGAAAVL RRDTRFTHW KNHLNRSVMG TGAMLLLFYA VTHLPLTTGV  
 101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPSFRSGQE  
 151 PAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSAT GVAMSSVWAT  
 201 LTGWHLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV  
 251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPIAFK QRLQALFRQR  
 301 \*

15 ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

20 orf135ng-1.pep MDTAKKDILGSGWMLVAAACFTVMNVLIKEASAKFALGSGELVFWRMFLSTVTLGAAAVL  
 orf135-1 MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL  
 20 orf135ng-1.pep RRDTRFTHWKNHLNRSVMGTGAMLLLFYAVTHLPLTTGVTLSTSSIFLAVFSFLILKE  
 orf135-1 RRDTRFTHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE  
 25 orf135ng-1.pep RISVYTQAVLLLLGFAGVLLLNPSFRSGQEPALAGLAGGAMSGWAYLKVRELSLAGEPG  
 orf135-1 RISVYTQAVLLLLGFAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG  
 30 orf135ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHLSFSAVYLSGIGVSALIAQLSMTRAYKVGDKFT  
 orf135-1 WRVVFYLSVTVAMSSVWATLTGWHLSFSAVYLSGIGVSALIAQLSMTRAYKVGDKFT  
 orf135ng-1.pep VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPIAFKQRLQALFRQR  
 orf135-1 VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPIAFKQRLQSLFRQR

35 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 66

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 549>:

40 1 ATGAAGCGGC GTATAGCCGT CTTGCTCCTG TTCCCGCAGA TAATCCGAGT  
 51 TTTGGGACAA CTGTTGCCGA AAATCGTCAA TACAGTTCCG GCACATCGGA  
 101 TGCTCTTCCA GATTTTCGGG ATGTTCTTTT TCTTCATACA CCAGCAATAT  
 151 CTGCCCGGGA TCGCCGAAAT CGATTCCCA TCGCGCATCG TGTTCCGTGC  
 201 GCTCCTCTTC CGTCATCTGC CCGCGCATTG CCTGTATGGT AAAGCCGCCG  
 45 251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGTGATGT CGTCAACCGG  
 301 AACGCAACG cTTTCGCCTT GTTCGACATT GGTGAGTTCG CCsGGTTCAT  
 351 TGTTGAGCAC ACCGTAAATA TAAAGACCGT CAAAATAAAT ATCGTCGATC  
 401 CACATATGTT CGCAAATTTT GCCGTCTTCG CCGTCTTGGA AAAAAGGGAC  
 451 TTTGACCATG GCAAAATCCA AGGCGGAAAT AATGCGGCGG CGTTCCCAA  
 50 501 AAAGcTCGCG CCAAAATATAT TTGAATGTTT TACGGGCGCG TTCGTCGGCA  
 551 CGGTTTACCG GTTCGCTGCG CTGTTCTACA TAATAAATGA CGGAATCGCC  
 601 CATCATATCT GCTCCTCAAC GTGTACGGTA TCTGTTTGCA CCTTACTGCG  
 651 GCTTTCTgCc kTCGGCATCC GATTCCGATT TGAAAAGTTC mmrwyATTG  
 701 GAATAG

55 This corresponds to the amino acid sequence <SEQ ID 550; ORF136>:

1 MKRRIAVFVL FPQIIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFIHQQY  
 51 LPGIAEIDSP CGIVFGALLF RHLPAHCLYG KAAVGDAVAH EHPVADVNNR

-320-

101 NANAFALFDI GQFAXFIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD  
 151 FDHGKIQQGN NAAAFPKKLA PKIFECFTGA FVGTVYRFVC LFYIINDGIA  
 201 HHSAPQVRVY LFAPYCGFLP SASDSLKSS XXSE\*

Further work revealed the complete nucleotide sequence <SEQ ID 551>:

5           1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG  
           51 AGTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC  
         101 GGATGCTCTT CCAGATTTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA  
         151 TATCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCCG  
         201 TGGCTCCTC TTCCGTCATC TGCCCGCGCA TTGCCTGTAT GGTAAAGCCG  
 10          251 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC  
         301 CGGAACGCAA ACGCTTTCGC CTTGTTTCGAC ATTGGTCAGT TCGCCGGGTT  
         351 CATTGTTCAG CACACCGTAA ATATAAGAC CGTCAAAATA AATATCGTCG  
         401 ATCCACATAT GTTCGCAAAAT TTCGCCGCTC TCGCCGCTCT GGAAGAAAGG  
         451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCC  
 15          501 AAAAAAGCTC GCGCCAAAAA TATTGAATG TTTTACGGGC GCGTTCGTCG  
         551 GCACGGTTTA CCGGTTTCGTC TGCCTGTTCT ACATAATAAA TGACGGAATC  
         601 GCCCATCATT CTGCTCCTCA ACGGTACGG TATCTGTTT CACCTTACTG  
         651 CGGCTTCTG CTTTCGGCAT CCGATTCCGA TTTGAAAGT TCCAAATATT  
         701 CGGAATAG

20 This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

          1 MMKRRIAEV LFPQIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHOQ  
           51 YLPGIAEIDS PCGIVFGALL FRHLPAHCLY GKAAGDAVA HEHPVADVNN  
         101 RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR  
         151 DFDHGKIQQG NNAAFPKKL APKIFECFTG AFVGTVYRFV CLFYIINDGI  
 25          201 AHHSAPQVRV YLFAPYCGFL PSASDSLKS SKYSE\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N. meningitidis*:

30		10	20	30	40	50	59
	orf136.pep	MKRRIAEVFLFPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHOQYLP	GIAEIDS				
	orf136a	MMKRRIAEVFLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHOQYLP	GIAEIDS				
35		10	20	30	40	50	60
	orf136.pep	PCGIVFGALLFRHLPAHCLYGKAAVGDVAHEHPVADVNNRNANAFALFDIGQFAXFIVQ	YLP				
	orf136a	PCGIVFGTLLFRHXSTHCLYGKAAVGNVAHEHPVADVNNRNANAFALFDIGQFAGFIVQ	YLP				
40		70	80	90	100	110	120
	orf136.pep	HTVNIKTVKINIVDPHMFANFAVFAVLEKRD	FDHGKIQQGNNAAFPKKLAPKIFECFTG				
45		130	140	150	160	170	180
	orf136a	HAINVKTVKINIVDPHMFANFAVFAVLEKRAL	TMAKSKXXMRRSQKSSRQKYLNVLRA				
50		180	190	200	210	220	230
	orf136.pep	AFVGTVYRFVCLFYIINDGIAHH--SAPQVRVYLFAPYCGFLPSASDSLKSSXXSEX	YLP				
	orf136a	R---SPARFTGLSACSTXXMTESPIISAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX	YLP				

The complete length ORF136a nucleotide sequence <SEQ ID 553> is:

55           1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG  
           51 GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC  
         101 GGATGCTCTT CCAGATNTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA  
         151 TACCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCCG  
         201 TACGCTCCTC TTCCGTCATC NGTCCACGCA TTGCCTGTAT GGTAAAGCCG  
         251 CCGTAGGGAA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC

-321-

5  
301 CGGAACGCAA ACGCTTTCGC CTTGTTCGAC ATTGGTCAGT TCGCCGGGTT  
351 CATTGTTCAG CACGCCATAA ATGTAAAGAC CGTCAAAATA AATATCGTCG  
401 ATCCACATAT GTTCGCAAAAT TTCGCCNTCT TCGCCGTCTT GGAAAAAAGG  
451 GCTTTGACCA TGGCAAAATC TAAGGNGNNA NNGATGCGGC GGCGTTCCCA  
501 AAAAAGCTCG CGCCAAAAAT ATTTGAATGT TTTGCGGGCG CGTTCGCCCG  
551 CACGGTTTAC CGGTTTGTCT GCCTGTTCTA CATAATAAAT GACGGAATCG  
601 CCCATCATAT CTGCTCCTCA ACGGTACGG TATCTGTTG CACCTTACTG  
651 CGGCTTCTG CCTTCGGCAT CCGATTCCGA TTTGAAAAGT TCCAAATATT  
701 CGGAATAG

10 This encodes a protein having amino acid sequence <SEQ ID 554>:

1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFIHHQ  
51 YLPGIAEIDS PCGIVFGTLL FRHXSTHCLY GKAAVGNVA HEHPVADV  
101 RNANAFALFD IGQFAGFIVQ HAINVKTVKI NIVDPHMFAN FAXFAVLEKR  
151 ALTMASKXXX XMRRRSQKSS RQKYLNVLRA RSPARETGLS ACST\*\*MTES  
201 PIISAPQVRV YLFAPYCGFL PSASDSLKS SKYSE\*

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

20  
25  
30  
35  
40

orf136a.pep	10	20	30	40	50	60
orf136-1	10	20	30	40	50	60
orf136a.pep	70	80	90	100	110	120
orf136-1	70	80	90	100	110	120
orf136a.pep	130	140	150	160	170	180
orf136-1	130	140	150	160	170	180
orf136a.pep	190	200	210	220	230	
orf136-1	190	200	210	220	230	

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from *N.gonorrhoeae*:

45  
50  
55

orf136.pep	59
orf136ng	60
orf136.pep	119
orf136ng	120
orf136.pep	179
orf136ng	180
orf136.pep	234
orf136ng	235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

60  
1 ATGATGAAGC GGCGTATAGC CGTCTCGTC CTGCTCATGC AGAAAAATCCG  
51 GATTTTGGGA CAACTGTTGC CGAAATCGT CAATACAGTT CCGGCACATC

10

This encodes a protein having amino acid sequence <SEQ ID 556>:

15

20

ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

25

30

35

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 67

40

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 557>:

45

50

This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

```

1  MENMVTFSKI RPLLAIAAAA LLAXRTAGN NAVRKPVQTA KPAAVVGLAL
51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGNLF ASGMSPDRLI
101 LEAEILGKTD LVDLTLSNG FIKGAKLONY INRKLGRMOI QOFFIKFAA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 559>:

-323-

1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC  
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GGCGGGAAAT AATGCTGTCC  
 101 GCAAGCCGGT GCAAAACGCC AAACCCGCG CAGTGGTCGG TTTGGCACTC  
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT  
 5 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
 251 CGATTGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA  
 301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC  
 351 CACCAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT  
 10 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AGGGGAATGC  
 501 CGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCAATGTG TTCCAACCCG  
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG  
 601 CCCGTCAGTG CCGCCGCGCG GCAGGGGCGG AATTTCTGTA TTGCCGTCGA  
 651 TATTTCCGCC CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC  
 15 701 TCGATCAGAG GCTGAACGTA ATGAGCGTTT CTGCGTTGCA AAATGAGTG  
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT  
 801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCAG  
 851 CACGTGCCGC ATTGCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
 901 TGA

20 This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL  
 51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSFDRLE  
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV  
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV  
 201 PVSAARRQGA NFVIAVDISA RPKNISQGF FSYLDQTLNV MSVSALQNEL  
 251 QQADVVIKPK VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY  
 301 \*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf137.pep		MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTA	KPAAVVGLALGGGASKGFAH				
35	orf137a	MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTA	KPAAVVGLALGGGASKGFAH				
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf137.pep	VGIIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSFDRLE	LEAEILGKTDLVDLTLSTNG				
	orf137a	VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSFDRLE	LEAEILGKTDLVDLTLSTSG				
		70	80	90	100	110	120
45	orf137.pep	FIKGAKLQNYINRKLGRMQIQQFPIKFAA					
	orf137a	FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV					
		130	140	149			
		130	140	150	160	170	180

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

50 1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC  
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GGCGGGAAAT AATGCTGCCC  
 101 GCAAGCCGGT GCAAAACGCC AAACCCGCG CAGTGGTCGG TTTGGCACTC  
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT  
 55 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
 251 CGATAGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA  
 301 TTGGAAGCCG AAATTTTAGG TAAAACCGAT TTGGTCGATT TAACCTTGTC  
 351 CACCAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT  
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGGAATGC  
 60 501 CGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCAATGTG TTCCAACCCG  
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG

5  
 601 CCCGTCAGTG CCGCCCGGCG GCANGNNNG NATNTCGTGA TTGCCGTCGA  
 651 TATTTCCGCC CGTCCGAGCA AAAACATCAG CCAAGGCTTC TTCTCTTATC  
 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CCGCGTTGCA AAATGAGTTG  
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT  
 801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCAG  
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
 901 TGA

This encodes a protein having amino acid sequence <SEQ ID 562>:

10  
 1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVGLAL  
 51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE  
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRVKGGRI QQFPIKFAAV  
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV  
 201 PVSAARRXXX XXVIAVDISA RPSKNISQGF FSYLDQTLNV MSVSALQNEL  
 251 GQADVVIKPQ VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAYRY  
 15 301 \*

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

20  
 orf137a.pep MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH  
 orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH  
 orf137a.pep VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG  
 orf137-1 VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG  
 25 orf137a.pep FIKGEKLQNYINRVKGGRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
 orf137-1 FIKGEKLQNYINRVKGGRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
 30 orf137a.pep FQPVIIGRHTYVDGGLSQPVVSAARRXXXXXVIAVDISARPSKNISQGFFSYLDQTLNV  
 orf137-1 FQPVIIGRHTYVDGGLSQPVVSAARRQGANFVIAVDISARPGKNISQGFFSYLDQTLNV  
 orf137a.pep MSVSALQNELGQADVVIKPQVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY  
 35 orf137-1 MSVSALQNELGQADVVIKPQVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from *N.gonorrhoeae*:

40  
 orf137.pep MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH 60  
 orf137ng MENMVTFSKIRSFLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH 60  
 45 orf137.pep VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLVDLTLSTNG 120  
 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMSPDRLELEAEILGKTDLVDLTLSTSG 120  
 orf137.pep FIKGAKLQNYINRKLGMQIQFPIKFAA 149  
 50 orf137ng FIKGEKLQNYINRVKGGRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV 180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

55  
 1 ATGGAATA TGTAACGTT TTCAAAATC AGATCATTTT TGGCAATCGC  
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGTAC GGCGGAAAC AATGCCGCC  
 101 GCAAGCCGGT GCAAACCGCC AAACCGCCG CAGTGGTCGC TTGGCACTC  
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT ATAGGAATTG TTAAGGTTT  
 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGT  
 251 CGATAGTCGG CAGCCTTTTG GCATCGGGTA TGTCGCCCCG CCGCCTCGAA  
 301 TTGGAAGCCG AGATTTTAGG TAAAACCGAT TTAGTCGATT TAACCTTGTC  
 351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
 60 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCGGT  
 451 GCCACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGAATGC



-325-

501 CGGGCAGGCG GTTCGTGCTT CCGCCGCCAT TCCCAATGTG TTCCAGCCAG  
 551 TCATCATCGG CAGGCACAAA TATGTTGACG GCGGTCTGTC GCAGCCCGTG  
 601 CCCGTCAGTG CCGCTCGGCG GCAGGGGCG AATTTCTGTA TTGCCGTCGA  
 651 TATTTCCGCA CGTCCGAGCA AAAATGTGCG TCAAGGTTTC TTCTCTTATC  
 701 TCGATCAGAC GCTGAACGTG ATGAGCGTTT CCGTGTTGCA AAACGAGTTG  
 751 gggcAGGCGG ATGTGGTTAT CAAACCGCag gtTTTGATT TGGGTGCAGT  
 801 CGGCGGATTC GATCAGAAAA AGCGCGCCAT CCGTTGGGC GAGGAGGCAG  
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
 901 TGA

10 This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVTFSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL  
 51 GGGASKGFAH IGIVKVLKEN GIPVKVVTGT SAGSIVGSL ASGMSFDRLE  
 101 LEAEILGKTD LVDLTSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV  
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHK YVDGGLSQPV  
 15 201 PVSAARRQGA NFVIAVDISA RPSKNVGQGF FSYLDQTLNV MSVSVLQNEL  
 251 GQADVVIKPO VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY  
 301 \*

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20 orf137ng MENMVTFSKIRSFLAIAAAA LLAACGTAGNNAARKPVQTA KPAAVVALALGGGASKGFAH  
 orf137-1 MENMVTFSKIRFLLAIAAAA LLAACGTAGNNAVRKPVQTA KPAAVVGLALGGGASKGFAH  
 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMSFDRLELEAEILGKTDLVDLTSTSG  
 25 orf137-1 VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSFDRLELEAEILGKTDLVDLTSTSG  
 orf137ng FIKGEKLQNYINRKVGGRQIQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
 orf137-1 FIKGEKLQNYINRKVGGRQIQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
 30 orf137ng FQPVIIGRHKYVDGGLSQPVVSAARRQGANFVIAVDISARPSKNVGQGF FSYLDQTLNV  
 orf137-1 FQPVIIGRHTYVDGGLSQPVVSAARRQGANFVIAVDISARPGKNISQGF FSYLDQTLNV  
 35 orf137ng MSVSVLQNELGQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY  
 orf137 MSVSALQNELGQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and  
 40 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

45 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA  
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCTGcTG CCGCTTTCCT  
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA  
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCmAT ATGCGGCAGG CCGGTTTGAA  
 201 CCGCGACCCC AAAACGGTCA AAGCCGTTTT TGCGGAAACG GCAAAAGGCG  
 251 GTTTGGAAC TCCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA  
 50 301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA  
 351 ACACGAAGG CTGCTATTC..

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL PLSCLHTLGN RLGHLAFYLL  
 51 KEDRARIVAX MRQAGLNPD KTVKAVFAET AKGGLELAPA FFRKPEDIET  
 101 MFKAVHGWEH VQALDKHEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
    51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCCTGCTG  CCGCTTTCCT
  101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
    5  151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGGCAGG  CGGGTTTGAA
    201  CCCCACCCC  AAAACGGTCA  AAGCCGTTTT  TCGCGAAACG  GCAAAAGGCG
    251  GTTTGGAAct  TGCCCCCGCG  TTTTTCAGAA  AACCGBAAGA  CATAGAAACA
    301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGGACAA
   10  351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTGG
    401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCCTGAC  CGCCATGTAC
    451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
    501  TCGCGGCAAA  GGAAAAACCG  CGCCTACCAG  CATACAAGGG  GTCAAACAAA
    551  TCATCAAAGC  CCTGCGTTCG  GGCGAAGCAA  CCATCGTCCT  GCCCACCAC
    601  GTCCCCTCCC  CTCAAGAGG  CGGGAAGGC  GTATGGGTGG  ATTTCTTCGG
   15  651  CAAACCTGCC  TATACCATGA  CGCTGGCGGC  AAAATTGGCA  CACGTCAAAG
    701  CCGTGAAAAC  CCTGTTTTTC  TGCTGCGAAC  GCCTGCCTGG  CGGACAAGGT
    751  TTCGATTTGC  ACATCCGCC  CGTCCAAGGG  GAATTGAACG  GCGACAAAGC
    801  CCATGATGCC  GCCGTGTTC  ACCGCAATGC  CGAATATTGG  ATACGCCGTT
    851  TTCCGACGCA  GTATCTGTTT  ATGTACAACC  GCTACAAAAT  GCCGTAA

```

20 This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

      1  MFRLQFRLFP  PLRTAMHILL  TALLKCLSL  LPLSCLHTL  GNRLGHLAF  YLLKEDRARI  VAX
    51  KEDRARIVAN  MRQAGLNPDF  KTVKAVFAET  AKGGLELAPA  FFRKPEDLET  MFKAVHGW  EHVVQALDK  HEG
  101  MFKAVHGW  EHVVQALDK  HEGLLFITP  HIGSYDLGG  RYISQLPFP  LITAMY
    25  151  KPPKIKAI  DKIMQAGRV  RGKGTAPT  SIQGVKQII  KALRSGEAT  IVLPDH
    201  VPSPQEGG  EGWVVDFF  GKPYMTLA  AKLHVKG  VKTLFFCC  ERLPGGQG
    251  FDLHIRPV  QGELNGDK  AHDAVF  FNRNAYW  IRRFPTQ  YLFMYNRY  KMP*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

30 *meningitidis*:

```

      10      20      30      40      50      60
orfl38.pep  MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX
      |||
orfl38a     MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
      10      20      30      40      50      60

      70      80      90     100     110     120
orfl38.pep  MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDLETMFKAVHGWHEVVQALDKHEG
      |||
orfl38a     MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDLETMFKAVHGWHEVVQALDKHEG
      70      80      90     100     110     120

orfl38.pep  LLF
orfl38a     LLFITPHIGSYDLGGRYISQQLPFPPLTAMYKPPKIKAI  DKIMQAGRV  RGKGTAPT  SIQGVKQII  KALRSGEAT  IVLPDH
      130     140     150     160     170     180

```

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
    50  51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCCTGCTG  CCGCTTTCCT
   101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
    151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGTCAGG  CAGGCATGAA
    201  TCCCACCCC  AAAACGGTCA  AAGCCGTTTT  TCGCGAAACG  GCAAAAGGCG
    251  GTTTGGAAct  TGCCCCCGCG  TTTTTCAGAA  AACCGBAAGA  CATAGAAACA
    301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGGACAA
    351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTGG
    401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCCTGAC  CGCCATGTAC
    451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
    501  TCGCGGCAAA  GGAAAAACCG  CGCCTACCAG  CATACAAGGG  GTCAAACAAA
    551  TCATCAAAGC  CCTGCGTTCG  GGCGAAGCAA  CCATCGTCCT  GCCCACCAC

```

-327-

5  
 601 GTCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG  
 651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG  
 701 GCGTGAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT  
 751 TTCGATTGCG ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC  
 801 CCATGATGCC GCGGTGTTCA ACCGCAATGC CGAATATGG ATACGCCGTT  
 851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAAT GCCGTAA

This encodes a protein having amino acid sequence <SEQ ID 570>:

10  
 1 MFRLQFRLFP PLRTAMHILL TALLKCLSL LPLSCLHTLGN RLGHIAFYLL  
 51 KEDRARIVAN MRQAGLNPD P KTVKAVFAET AKGGLLELAPA FFRKPEDIE T  
 101 MFKAVHGW EHVQQALDKHEG LLFITPHIGS YDLGGYISQ QLPFPLTAMY  
 151 KPPKIKAI D KIMQAGRVRG KGTAPT SIQG VKQIIKALRS GEATIVLPD H  
 201 VPSPQEGG EG VWVDFFGKPA YTM TLA AKLA HVKGVKTLFF CCE RLPGGQG  
 251 FDLHIRPVQ G ELNGDKAHD A AVFN RNAEY W IRRFPTQYLF MYNRYKMP\*

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

15  
 orf138a.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHIAFYLLKEDRARIVAN  
 orf138-1 MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHIAFYLLKEDRARIVAN  
 20  
 orf138a.pep MRQAGMNPDPKTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGW EHVQQALDKHEG  
 orf138-1 MRQAGLNPD P KTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGW EHVQQALDKHEG  
 orf138a.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAI D KIMQAGRVRG KGTAPT SIQG  
 25  
 orf138-1 LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAI D KIMQAGRVRG KGTAPT SIQG  
 orf138a.pep VKQIIKALRSGEATIVLPD HVPSPQEGGEGVWVDFFGKPAYTMTLA AKLAHVKGVKTLFF  
 orf138-1 VKQIIKALRSGEATIVLPD HVPSPQEGGEGVWVDFFGKPAYTMTLA AKLAHVKGVKTLFF  
 30  
 orf138a.pep CCE RLPGGQGF D LHIRPVQ G ELNGDKAHD A AVFN RNAEY W IRRFPTQYLFMYNRYKMP  
 orf138-1 CCE RLPGGQGF D LHIRPVQ G ELNGDKAHD A AVFN RNAEY W IRRFPTQYLFMYNRYKMP

### 35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from *N.gonorrhoeae*:

40  
 orf138.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHIAFYLLKEDRARIVAX 60  
 orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHIAFYLLKEDRARIVAN 60  
 orf138.pep MRQAGLNPD P KTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGW EHVQQALDKHEG 120  
 orf138ng MRQAGLNPD P KTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGW EHVQQALDKHEG 120  
 45  
 orf138.pep LLF 123  
 orf138ng LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAI D KIMQAGRVRG KGTAPTGIQG 180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

50  
 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CTTTTCGGA CCGCCATGCA  
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG TCGCTTTCCT  
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA  
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA  
 201 CCGGACACG CAGACGGTCA AAGCCGTTT TCGGGAACG GCAAAATGCG  
 55  
 251 GTTTGGAAC TGCCTCCGCG TTTTCAAAA AACCGGAAGA CATCGAAACA  
 301 ATGTTCAAAG CGGTACACGG CTGGGAACAC GTGCAGCAGG CTTTGGACAA  
 351 GGGCGAAGGG CTGCTGTTCA TCACGCCGCA CATCGGCAGC TACGATTGCG  
 401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCACCTGAC CGCCATGTAC  
 451 AAGCCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT  
 60  
 501 GCGCGGCAAA GGCAAAACcg cgccaccgg catACAAGGG GTCAAAACAA  
 551 tcatcaAGGC CCTGCGCGCG GCGGAGGCAA CCATcATCCT GCGGACAC

-328-

5 601 GTCCCTTCTC CGCAGGAagg cggCGGCGTG TGGGCGGATT TTTTCGGCAA  
 651 ACCTGCATAc accCATGACAC TGGCGGCAAA ATTGGCACAC GTCAAAGGCG  
 701 TGAAAACCCCT GTTTTCTGCG TCGGAACGCC TGCCCGACGG ACAAGGCTTC  
 751 GTGTTGCACA TCCGCCCCGT CCAAGGGGAA TTGAACGGCA ACAAAGCCCCA  
 801 CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTTTC  
 851 CGACGCAGTA TCTGTTTATG TACAACCGCT ATAAACGCC GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

10 1 MFRLQFRLFP PLRTAMHILL TALLKCLSL LSLCLHTLGN RLGHLAFYLL  
 51 KEDRARIVAN MRQAGLNPD TQTVKAVFAET AKCGLELAPA FFKKPEDTET  
 101 MFKAVHGWEH VQALDKGEG LLFITPHIGS YDLGGYISQ QLPFHLTAMY  
 151 KPPKIKAIK IMQAGRVRGK GKTAPTGIQG VKQIIKALRA GEATIILPDH  
 201 VPSPQEGGGV WADFFGKPAY TMTLAAKLAH VKGVKTLFFC CERLPDGGGF  
 251 VLHIRPVQGE LNGNKAHDAA VFNRNTEYWI RRFPTQYLFM YNRYKTP\*

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

15 orf138-1.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSLCLHTLGNRLGHLAFYLLKEDRARIVAN  
 orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSLSLSLCLHTLGNRLGHLAFYLLKEDRARIVAN  
 20 orf138-1.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDTETMFKAVHGWEHVQALDKHEG  
 orf138ng MRQAGLNPDQTQTVKAVFAETAKGGLELAPAFFRKPEDTETMFKAVHGWEHVQALDKHEG  
 orf138-1.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQG  
 25 orf138ng LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQG  
 orf138-1.pep VKQIIKALRSGEATIVLPDHPVSPQEGGGVWVDFFGKPAYTMTLAAKLAHVKGKTLFF  
 orf138ng VKQIIKALRAGEATIILPDHPVSPQEGG-GVWADFFGKPAYTMTLAAKLAHVKGKTLFF  
 30 orf138-1.pep CCERLPGGQGFDLHIRPVQGE LNGDKAHDAAVFNRNTEYWIRRFPTQYLFMYNRYKMP  
 orf138ng CCERLPDGGQFVLHIRPVQGE LNGNKAHDAAVFNRNTEYWIRRFPTQYLFMYNRYKTP

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

35 gnl|PID|e334283 (Y14568) htrB [Pseudomonas fluorescens] Length = 253  
 Score = 80.8 bits (196), Expect = 9e-15  
 Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)  
 40 Query: 101 MFKAVHGWEHVQALDKGEGLLFITPHIGSYD-LGGYISQQLPFHLTAMYKPPKIKAIK 159  
 + + V G E +++AL G+G++ IT H+G+++ L Y SQ P Y+PPK+KA+D  
 Sbjct: 94 LVREVEGLEVLKEALASGKGVGITSHLGNWEVLNHFYCSQCKPI---IFYRPPKLAVD 150  
 Query: 160 KIMQAGRVRGKGKTAPTGIQGKQIIKALRAGEATIILPDHPVSPQEGGGVWADFFGKPA 219  
 ++++ RV+ K A + +G+ +IK +R G I D P P E G++ FF A  
 45 Sbjct: 151 ELLRKQVRVQLGNKVAASSTKEGILSVIKEVRKGGQVGIPAD--PEPAESAGIFVPFFATQA 208  
 Query: 220 YMTLAAKLAHVKGKTLFFCCERLPDGGGF 250  
 T + +F RLPDG G+  
 Sbjct: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239

50 Based on this analysis, including the presence of a putative transmembrane domain in the  
 gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
 their epi'opes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.  
 The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A  
 55 shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein  
 was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis

(Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

5      1  ..GCGTGGTCGG CCGGCGAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
      51  GCATGCGGTG TGGAAATACTT TGCCTTCTC GCGGCGGCG GTGTATGCGG
     101  CAGCGGTTTT GGGTGTGGTG TATGCGGCGC CGGCGGCGG CTCGCGCTGG
     151  ATGCGCGGGC TGATGTTTTA GCCGTTTATG GTGTGCGCGG TTTGTGTTTC
     201  GCGGCGCGTG CTGCTGCTTT ATCCGCGATG GACGGCTTCG TTGCCGTGTC
10     251  TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTGTGGC AAAAGATGTT
     301  TTATCAGCCT GGGATGCACT GCCGCCGAT TACGGCAGG CGGCGGCGGG
     351  TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCAC TCCCCCTCT
     401  TGAAACCGSC GTTGCAGGCG GGTCTGACTT TGGCGGCGGC AACCTGCGTG
     451  GCGGAATTTG CCGCGACATT GTTCTGTGCG CGTCCGGAAT GGCAGACGCT
15     501  GACGACTTTG ATTTATGCCT ATTTGGGACG CGCGGGTGAG GATAATTACG
     551  CGCGGCGGAT GGTGCTG..

```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

      1  ..AWSAGESWRV LMESETHAV WNTLRFSAVA VYAAAVLGTV YAAPARRSAW
     51  MRGLMFXPFM VSPVVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV
20    101  LSAWDALPPD YGRAAGLGA NGFQTACRIT FPLKLPALRR GLTLAAATCV
     151  GEFAATLFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL..

```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```

      1  ATGGATGGAC GCGTGTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCCTTC
     51  GGCTTTTGTG GCGGTAATGG TCGTTGCGCC TTGTGTGGCG GTGGCGGCGT
25    101  ATGACGGTTC GCGGTGGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAA
     151  CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT
     201  GCCTTTGGGC GTGCCTGTGCG CGTGGGTGCT GCGCGGCGTG GCGTTTCCGG
     251  GCGGGGCTTT GGTGTGCGC CTGCTGATGC TGCTTTTGTG GATGCCCACG
     301  TTGGTGGCGG GCGTGGGCGT GCTGGCCCTG TTCGGGGCGG ACGGGCTGTT
30    351  GTGGCGCGGC AGGCAGGATA CGCCGTATCT GTTGTGTAC GGCAATGTGT
     401  TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTGTGCAA
     451  GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CGGGGGCGTG
     501  GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCC TGGCTTGCCG
     551  GCGGCGGTG CTTGTCTTTT CTGTATTGTT TTTCCGGGTT CGGGCTGGCG
35    601  CTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA
     651  GTTGGTCATG TTCGAACCTG ATATGGCGGT TGCTTCGGTG CTGGTGTGGC
     701  TGGTGTGGG GGTAAACGGC GCGGCAGGGT TGCTGTATGC GTGGTTCGGC
     751  AGGCGCGCGG TTTCGGATAA GCGGTTTCC CCGTGTATGC CGTCGCGCGC
40    801  GCAGTCGGTC GGGGAATATG TGCTGCTGGC GTTGTGCGCG GCGGTGTTGT
     851  CTGTGTGCTG CCGTGTTCCT TTGTGGCAA TTGTGTGAA AGCGTGGTCG
     901  GCGGCGGAAT CGTGGCGTGT GTTAATGGAA AGTGAAACGT GGCAGGCGGT
     951  GTGGAATACT TTGCGCTTCT CGGCGGCGGC GGTGTATGCG GCGGCGGTTT
100   1001  TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGTGCGCGTG GATGCGCGGG
     1051  CTGATGTTTT TGCCGTTTTAT GGTGTGCGCG GTTGTGTTT CGGCGGGCGT
45   1101  GCTGCTGCTT TATCCGCAGT GGACGGCTTC GTTGCCGTTG CTGCTGGCGA
     1151  TGATATGCGT GCTGGCGTAT CCGTTGTGG CAAAAGATGT TTTATCAGCC
     1201  TGGGATGCAC TGCCGCCGGA TTACGGCAGG GCGGCGGCGG GTTGGGTGTC
     1251  AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCCCTC TTGAAACCGG
     1301  CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT
50   1351  GCGGCGACAT TGTTCCTGTC GCGTCCGGA TGGCAGACGC TGACGACTTT
     1401  GATTTATGCC TATTTGGGAC GCGCGGGTGA GGATAATTAC GCGCGGGCGA
     1451  TGGTGCTGAC ATTGCTGTTG GCGGCGTTCG CGCTGGGTAT TTCTCTGCTG
     1501  TTGGACGGCG GCGAAGGCGG AAAACAGACG GAAACGTTAT AA

```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

```

55    1  MDGRRVVWVG AFALLPSAFL AVMVVAPLWA VAAVDGLAWR AVLSDAYMLK
     51  RLAWTVFQAA ATCVLVLP LG VPVAVVLARL AFPGRALVLR LLMLPFVMP
100   101  LVAGVGVLAL FGADGLLWRG RQDTPYLLLY GNVFFNLPVL VRAAYQGFVQ
     151  VPAARLQTAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA

```

Computer analysis of this amino acid sequence gave the following results:

10 ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of *N. meningitidis*:

35 The complete length ORF139a nucleotide sequence <SEQ ID 577> is:

	1	ATGGATGGAC	GCGCTTGGGC	GGTATGGGGT	GCTTTTGCCC	TGCTGCCTTC
	51	GGCTTTTTTG	GCGGCAATGG	TCGTTGCGCC	TTTGTGGGCG	GTGGCGGCGT
	101	ATGACGGTTT	GGCGTGGCGC	GCGGTGCTGT	CGGATGCCTA	TATGCTCAAA
40	151	CGTTTGGCGT	GGACGGTATT	TCAGGCAGCG	GCAACCTGTG	TGTCGTGTCT
	201	GCCTTTGGGC	GTGCCCTGTCG	CGTGGGTGCT	GGCGCGGCTG	GCGTTTCCGG
	251	GGCGGGCTTT	GGTGCTGCGC	CTGCTGATGC	TGCTTTTGTG	GATGCCACG
	301	TTGGTGGCGG	GCGTGGGCGT	GCTGGCTCTG	TTCCGGGCGG	ACGGCCTGTN
	351	GTGGCGCGCG	TGGCAGGATA	CGCCGTATCT	GTTGTTGTAC	GGCAATGTGT
45	401	TTTTTNACCT	TCCTGTGTTG	GTCAGGGCGG	CATATCAGGG	GTTTGTGCAA
	451	GTGCCTGCGG	CACGGCTTCA	GACGGCACNG	ACATTGGGCG	CGGGGGCGTG
	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCGT	TTTGGCGCCC	TGGCTTGGCG
	551	GCGGCGGTGT	CCTTGTCTTC	CTGTATTGTT	TTTCCGGGTT	GCGGCTGGCA
	601	TTGCTGCTGG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
50	651	GTTGGTCATG	TTCCGAACCT	ATATGGCCGT	TGCTTCGGTG	CTNGTGTGGC
	701	TGGTGTNNGG	GGTAACNCGC	GCGGCAGGGT	TGCTGTATGC	GTGGTTCCGC
	751	AGGCGCGCGG	TTTTCGGATA	GCGNGTTTCC	TGCTGTATGC	CTGCGCCGCC
	801	GCAGTCGGTC	GGGGAATATG	TGCTNCTGGC	GTTTGCGGCG	GCGGTGTNGT
	851	CTGTGTGCTG	CCTGTTTCNT	TTGTTGGCAA	TTGTTGTGAA	AGCGTGGTCG
55	901	GCCGGCGAAT	CGTGCGGTGT	GTTAATGGAA	AGTGAACGTT	GCGAGGCGGT
	951	GTGGAATACT	NTGGCGTCTT	CGCGCGCGCG	GCGTGTATCG	GCGCGGCTTT
	1001	TGGGTGTGGT	GTATGCGGCG	GCGGCGCGGC	GGTCGGCGTG	GATGCGCGGG
	1051	CTGATGTTTT	TGCCGTTTAT	GGTGTGCGCG	GTTTGTGTTT	CGGCGGGCGT
	1101	GCTGCTGCTT	NATCCGCAAT	GGACGGCTTC	GTTGCGCGTG	CTGCTGGCGA
60	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTTGTGG	CAAAAGATGT	TTTATCAGCC
	1201	TGNATGTCAC	TGCCGCCGGA	TTACGGCAGG	GCGGCGGCGG	GTTTGGGTGC
	1251	AAACGGCTTT	CAGACGGCAT	GCCGCATCAT	GTTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGCG	CGGTCGTACT	TTGGCGGCGG	CAACCTCGCT	GGGCGAATTT
	1351	GCGGCAACCT	TGTTTNTGTC	GCGTCNCGAG	TGGCAGACGC	TGACGACTTT

5	1	MDGRRWAVWG	AFALLPSAFL	AAMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVLPGL	VPVAWVLARL	AFPGRALVLR	LLMLPFVMP
	101	LVAGVGVLAL	FGADGLXWRG	WQDTPYLLLY	GNVFFXLPLV	VRAAYQGFVQ
	151	VPAARLQTAX	TLGAGAWRRF	WDIEMPVLPR	WLAGGVCLVF	LYCFSGFGLA
10	201	LLLGSGRYAT	VEVEIYQLVM	FELDMAVASV	LVLWLVXGVT	AAGLLYAWFG
	251	RRAVSDKAVS	PVMPSPPPQSV	GEYVLLAFAA	AVXSVCCFLX	LLAIIVKAWS
	301	AGESWRVLME	SETWQAVWNT	XRFSAAAVYA	AAVLGVVYAA	AAARRSAWMRG
	351	LMFLPFMVSP	VCVSAGVLLL	XPQWTASLPL	LLAMYALLAY	PFVAKDVLSA
15	401	XDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAATCVGEF
	451	AATLFXSRXE	WQTLTTLIYA	YXGRAGXDNY	ARAMVLTLLL	AAFALGXFLL
	501	LDGGEGGKRT	ETL*			

20 orf139a.pep MDGRRWVWVGAFALLPSAFLAAMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA  
 orf139-1 MDGRRWVWVGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA

25 orf139a.pep ATCVLVLPPLGVFPAWVRLARLAFFGRALVLRLLMLPFVMPMTLVAGVGVLALFGADGLXWRG  
 orf139-1 ATCVLVLPPLGVFPAWVRLARLAFFGRALVLRLLMLPFVMPMTLVAGVGVLALFGADGLLWRG

30 orf139a.pep WQDTPYLLLYGNVFFXLPVLVRAAYQGFVQVPAARLQTAXTLGAGAWRRFWDIEMPVLRP  
 orf139-1 RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP

35 orf139a.pep WLAGGVCLVFLYCFSGFGLALLLGGSRATVEVEIYQLVMFELDMAVASVLVWLVLXGVTA  
 orf139-1 WLAGGVCLVFLYCFSGFGLALLLGGSRATVEVEIYQLVMFELDMAVASVLVWLVLGVTA

40 orf139a.pep AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFAAAVXSVCCFLXLLAIIVVKAWS  
 orf139-1 AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFAAAVLSVCCFLPLLAIIVVKAWS

45 orf139a.pep AGESWRVLMESSETWQAVWNTXRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP  
 orf139-1 AGESWRVLMESSETWQAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP

50 orf139a.pep VCVSAGVLLLPQWTASLPLLLAMYALLAYPFVAKDVLSDAXDALPPDYGRAAGLGANGF  
 orf139-1 VCVSAGVLLLPQWTASLPLLLAMYALLAYPFVAKDVLSDAWDALPPDYGRAAGLGANGF

orf139a.pep QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFXSRXEWQTLTTLLIYAYXGRAGXDNY  
 orf139-1 QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLLIYAYLGRAGEDNY

orf139a.pep ARAMVLTLLLAALGXFLLLDGGEGGKRTETLX  
 orf139-1 ARAMVLTLLLAALGIFLLLDGGEGGKOTETLX

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from  
55 *N.gonorrhoeae*:

	orf139.pep	AWSAGESWRVLMSESETWHAVWNTLRFSAAA	30
		:	
	orf139ng	QSVGEYVLLAFSVAVLSVCCFLFLSAIVVKAWSAGESRRVLMSESETWQAVWNTLRFSAAA	327
60	orf139.pep	VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLYPQWTASLP LLAMYAL	90
		:	
	orf139ng	VFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSPVCVSAGVLLYPGWTASLP LLAMYAL	387

orf139.pep	LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV	150
orf139ng		447
5 orf139.pep	GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL	189
orf139ng		507

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

10	1	MDGRCWAVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVLP LG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMPT
	101	LVAGVGV LAL	FGADGLLWRG	RQDTPYLLLY	GNVFFNLPVL	VRAAYQGFAQ
	151	VPAARLQTAR	TLGAGAWRPF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
15	201	LLLGGSRYAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCCLFP	LSAIVVKAWS
	301	AGESRRVLME	SETWOAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVWMRG
	351	LVFLPFMVSP	VCVSAGVLLL	YPGWTASLPL	LLAMYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
20	501	LDN GEGGKRT	ETL*			

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:

	1	ATGGATGGAC	GGTGTGGGC	GGTACGGGGT	GCTTTTCCC	TGCTGCCTTC
	51	GGCTTTT	GGCGTAATGG	TCGTTGCGCC	TTTGTGGGCG	GTGGCGGCGT
25	101	ATGACGTTT	GGCGTGGCGC	GCGGTGCTGT	CGGATGCCTA	TATGCTCAAA
	151	CGTTTGGCGT	GGACGGTGT	TCAGGCGGCG	GCAACCTGTG	TGCTGGTGCT
	201	GCCTTTGGGC	GTGCCTGTCG	CGTGGGTGCT	GGCGCGGCTG	GCGTTCCCGG
	251	GGCGGGCTTT	GGTGCTGCGC	CTGCTGATGC	TGCCGTTTGT	GATGCCCACG
	301	CTGGTGGCGG	GCGTGGGCGT	GCTGGCTCTG	TTCCGGGCGG	ACGGGCTGTT
30	351	TGGGCGCGCG	CGGCAGGATA	CGCCGTATCT	GTTGTGTAC	GGCAATGTGT
	401	TTTTCAACCT	GCCCGTGTG	GTCAGGGCGG	CGTATCAGGG	GTTTGCTCAA
	451	GTGCCTGCGG	CACGGCTTCA	GACGGCACGG	ACGTTGGGCG	CGGGGGCGTG
	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCCG	TTTGCGCCCG	TGGCTTGCCG
	551	GCGGCGTG	CCTGTCTTTC	CTGTATTGTT	TTTCGGGGTT	CGGGCTGGCA
35	601	TTGCTGTTGG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
	651	GTTGGTTATG	TTCGAAC TCG	ATATGGCGGG	GGCTTCGGCG	CTGGTGTGGC
	701	TGGTGTGGG	GGTAACGGCG	GCGGCAGGGT	TGCTGTATGC	GTGGTTCGGC
	751	AGGCGCGCGG	TTTCGGATAA	GGCGGTTTCC	CCCGTGATGC	CGTCGCCGCC
	801	GCAATCGGTG	GGGGAATATG	TATTGCTGGC	ATTTTCGGTG	GCGGTGTTGT
40	851	CCGTGTGCTG	CCTGTTTCTT	TTGTGCGCAA	TTGTGTGAA	AGCGTGGTCG
	901	GCCGGCGAAT	CGCGGCGTGT	GTTAATGGAA	AGTGAACGCT	GGCAGGCAGT
	951	GTGGAATACT	ttGCGCTTTT	CGGCGGCGGC	GGTGTTTGCG	GCGGCGGTTT
	1001	TGGGTGTGGT	GTATGCGGCG	GCGGCGCGGC	GGCTGGTGTG	GATGCGCGGA
	1051	CTGGTGTTTT	TACCGTTTAT	GGTGTCGCCG	GTTTGTGTTT	CGGCGGGCGT
45	1101	GCTGCTGCTT	TATCCGGGGT	GGACGGCTTC	GTTACCGCTG	CTGCTGGCGA
	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTTGTGG	CAAAAGATGT	TTTATCGGCC
	1201	TGGGATGCAC	TGCCGCCGGA	TTACGGCAGG	GCGGCGGCAG	GTTTGGGCGC
	1251	AAACGGCTTT	CAGACGGCAT	GCCGTATCAC	GTTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGCG	CGGTCTGACT	TTGGCGGCGG	CGACGTGTGT	GGGCGAATTT
50	1351	GCGGCAACCT	TGTTCTCTGC	GCGTCCGGAA	TGGCAGACGT	TGACGACTTT
	1401	GATTTATGCC	TATTTGGGGC	GTGCGGGTGA	GGACAATTAT	GCGCGGGCAA
	1451	TGGTGTGAC	ATTGCTGTTG	TCGGCATTTC	CGGTGTGCAT	TTTCTGCTG
	1501	TTGACAACG	GCGAAGGCGg	aaaACGGACG	GAAACGTTAT	AA

This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-1>:

55	1	MDGRCWAVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVLP LG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMPT
	101	LVAGVGV LAL	FGADGLLWRG	RQDTPYLLLY	GNVFFNLPVL	VRAAYQGFAQ
	151	VPAARLQTAR	TLGAGAWRRF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
	201	LLLGGSRYAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
60	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCCLFP	LSAIVVKAWS
	301	AGESRRVLME	SETWOAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVWMRG
	351	LVFLPFMVSP	VCVSAGVLLL	YPGWTASLPL	LLAMYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
	501	LDN GEGGKRT	ETL*			



ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

```

5      orf139ng      MDGRCWAVRGAFSLLESAFLAVMVVAPLWAVAAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
      orf139-1      MDGRRWVVWGAFALLESAFLAVMVVAPLWAVAAAYDGLAWRAVLSDAYMLKRLAWTVFQAA

      orf139ng      ATCVLVLPPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPPTLVAGVGVLALFGADGLLWRG
      orf139-1      ATCVLVLPPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPPTLVAGVGVLALFGADGLLWRG

10     orf139ng      RQDTPYLLLYGNVFFNLPVLVRAAYQGFAQVPAARLQTARTLGAGAWRREWDIEMPVLRP
      orf139-1      RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLQTARTLGAGAWRREWDIEMPVLRP

      orf139ng      WLAGGVCLVFLYCFSGFGLALLGGSRyatVEVEIYQLVMFELDMAGASALVWLVLGVTA
      orf139-1      WLAGGVCLVFLYCFSGFGLALLGGSRyatVEVEIYQLVMFELDMAVASVLVWLVLGVTA

      orf139ng      AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFSVAVLSVCCLFPLSAIVVKAWS
      orf139-1      AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFSVAVLSVCCLFPLSAIVVKAWS

20     orf139ng      AGESRRVLMESSETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
      orf139-1      AGESRRVLMESSETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP

25     orf139ng      VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF
      orf139-1      VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF

30     orf139ng      QTACRITFPLLPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
      orf139-1      QTACRITFPLLPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY

      orf139ng      ARAMVLTLLLSAFAVCIFLLLDNGEGGKRTETL
35     orf139-1      ARAMVLTLLLSAFALGIFLLLDGGEGGKQTETL

```

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

```

45      1  ATGGACGGCT  GGACACAGAC  GCTGTCCGCG  CAAACCCTGT  TGGGCATTTC
      51  GCGGCGGCA  ATCATCCTCA  TTCTGATTTT  AATCGTCAGA  TTCCGCATCC
      101  ACGCGCTGCT  GACACTGGTC  ATCGTCAGCC  TGCTGACGGC  TTTGGCAACC
      151  GGTTCGCCA  CAGGCAGCAT  TGTCAAAGAC  ATACTGGTCA  AAAACTTCGG
      201  CGGCACGCTC  GCGGCGTGG  CGCTTCTGGT  CGGCCTGGGC  GCGATGCTCG
      251  AACGTTGGT  C...

```

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

```

50      1  MDGWTQTLTA  QTLGLISAAA  IILILILIVR  FRIHALTLV  IVSLLTALAT
      51  GLPTGSIVKD  ILVKNFGGTL  GGVALLVGLG  AMLERLV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

```

55      1  ATGGACGGCT  GGACACAGAC  GCTGTCCGCG  CAAACCCTGT  TGGGCATTTC
      51  GCGGCGGCA  ATCATCCTCA  TTCTGATTTT  AATCGTCAAA  TTCCGCATCC
      101  ACGCGCTGCT  GACACTGGTC  ATCGTCAGCC  TGCTGACGGC  TTTGGCAACC
      151  GGTTCGCCA  CAGGCAGCAT  TGTCAACGAC  ATACTGGTCA  AAAACTTCGG

```

-334-

201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG  
 251 GACGTTTGGT CGAAACATCC GGC GGCGCAC AGTCGCTGGC GGACGCGCTG  
 301 ATCCGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCGCTC  
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC  
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCCTGCC  
 501 GCGCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG  
 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT  
 601 AGCGGTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT  
 651 TCCCGAATG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC  
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG  
 751 ATTTTCTCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TGCGGACGAA ACCTGGGTTC AGACGGCAAA AATAATCGGT TCGACACCGA  
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAAA  
 901 CGCGGCGAAA GCGCGAGCG GTTGGAAGAA ACCGTGGACG GCGCACTCGC  
 951 CCGCGTCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCCGCG  
 1001 GCGTTTTCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TTCCCGTCTT TTTGGGCTGT TTCCTTGTCT CCTTGGCACT  
 1101 GCGTATCGCG CAAGGTTCGG CAACCGTCGC CCTGACCACC GCCGCCGCGC  
 1151 TGATGGCTCC TGCCGTTGCC GCCGCCGGCT TTACCGACTG GCAGCTCGCC  
 1201 TGTATCGTAT TGGCAACGGC GGCAGGTTTC GTCGGTTGCA GCCACTTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGTCT CTTGGACATG GACGTACCGA  
 1301 CCACGCTGAA AACCTGGACG GTCAACCAA CCCTCATCGC ACTCATCGCG  
 1351 TTTGCCCTGT CCGCACTGCT GTTCGCCATC GTCTGA

25 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQTLA QTLGLISAAA IILILILIVK FRIHALLTLV IVSLLTALAT  
 51 GLPTGSIVND ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL  
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP  
 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL  
 251 IFLNTGVSAL ISEKLVSADE TWVQTAKIIG STPIALLISV LVALFVLGRK  
 301 RGESESALEK TVDGLAPVC SVLITGAGG MFGGVLRASG IGKALADSMA  
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGCSEFNDSG FWLVGRLLDM DVPTTLKWT VNQTLIALIG  
 451 FALSALLFAI V\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N. meningitidis*:

40 orf140.pep MDGWTQTLAQTLLGLISAAAIIILILILIVRFRIHALLTLVIVSLLTALATGLPTGSIVKD  
 orf140a MDGWTQTLAQTLLGLISAAAIIILILILIVKFRHALLTLVIVSLLTALATGLPTGSIVND  
 45 orf140.pep ILVKNFGGTLGGVALLVGLGAMLERLV  
 orf140a VLVKNFGGTLGGVALLVGLGAMLERLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF  
 50

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGACAGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTC  
 51 GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC  
 101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC  
 151 GGTTCGCCCA CAGGCAGCAT TGTCAACGAC GACTGGTCA AAAACTTCGG  
 201 CGGCACGCTC GGCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG  
 251 GACGTTTGGT CGAAACATCC GGC GGCGCAC AGTCGCTGGC GGACGCGCTG  
 301 ATCCGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCGCTC  
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC  
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCCTGCC

-335-

5 501 GCCCATCCG GCGCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG  
 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT  
 601 AGCGGCTATA TGCTCGGCAA AGTGTTGGGG CGCAACCATCC ATGTTCCCGT  
 651 TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC  
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG  
 751 ATTTTCCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TCGCGACGAA ACCTGGGTTT AGACGGCAA AATAATCGGT TCGACACCGA  
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAAA  
 901 CGCGCGGAAA GCGGCAGCGC GTTGGAAAA ACCGTGGACG GCGCACTCGC  
 10 951 CCCCGTCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCCGGC  
 1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TTCCCGTCCT TTTGGGCTGT TTCCTTGTCG CCTTGGCACT  
 1101 GCGTATCGCG CAAGGTTCGG CAACCGTCGC CTTGACCACC GCCGCCGCGC  
 1151 TGATGGCTCC TGCCGTTGCC GCCGCCGGCT TTACCGACTG GCAGCTCGCC  
 1201 TGTATCGTAT TGGCAACGGC GGCAGGTTCC GTCGGTTGCA GCCACTTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGCCT CTTGGACATG GACGTACCGA  
 1301 CCACGCTGAA AACCTGGACG GTCAACCAA CCTCATCGC ACTCATCGGC  
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

This encodes a protein having amino acid sequence <SEQ ID 588>:

20 1 MDGWTQTLISA QTLGISAAAA IILILILIVK FRIHALLTLV IVSLLTALAT  
 51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL  
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVL  
 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPKAGTV VAIMLIPMLL  
 25 251 IFLNTGVSAL ISEKLVSADE TWVQTAKIIG STPIALLISV LVALFVLGRK  
 301 RGECSALEK TVDGLAPVC SVILITGAGG MFGGVLRSAG IGKALADSM  
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGCSEFNDG FWLVGRLLDM DVPTTLKWT VNQTLIALIG  
 451 FALSALLFAI V\*

30 ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

orf140-1.pep MDGWTQTLISAQTLGISAAAAIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60  
 orf140a MDGWTQTLISAQTLGISAAAAIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60  
 35 orf140-1.pep ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF 120  
 orf140a VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF 120  
 40 orf140-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLFPALASIGAFSVMHVFLPPHPGPIAASEFYG 180  
 orf140a GFPIFFDAGLIVMLPIVFATARRMKQDVLFPALASIGAFSVMHVFLPPHPGPIAASEFYG 810  
 orf140-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPVPELLSGGTQDNDLPKEPAKAGTV 240  
 45 orf140a ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPVPELLSGGTQDNDLPKEPAKAGTV 240  
 orf140-1.pep VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300  
 orf140a VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300  
 50 orf140-1.pep RGECSALEKTVDGALAPVCSVILITGAGGMFGGVLRSAGIGKALADSMADLGIPVLLGC 360  
 orf140a RGECSALEKTVDGALAPVCSVILITGAGGMFGGVLRSAGIGKALADSMADLGIPVLLGC 360  
 55 orf140-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEFNDG 420  
 orf140a FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEFNDG 420  
 60 orf140-1.pep FWLVGRLLDMDVPTTLKWTVNQTLIALIGFALSALLFAIV 461  
 orf140a FWLVGRLLDMDVPTTLKWTVNQTLIALIGFALSALLFAIV 461

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 *N.gonorrhoeae*:

```

orfl140.pep  MDGWTQTLQAQTLGLISAAAIIILILILIVFRIRALLTLVIVSLLTALATGLPTGSIVKD  60
              ||| |||||:|||||:|||||:|||||:
orfl140ng    MDGRTQTLQAQTLGLISAAAIIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND  60

5  orfl140.pep  ILVKNFGGTLGGVALLVGLGAMLERLV  87
              :|||||:
orfl140ng    VLVKNFGGTLGGVALLVGLGAMLGRVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF  120

```

The complete length ORF140ng nucleotide sequence <SEQ ID 589> was predicted to encode a protein having amino acid sequence <SEQ ID 590>:

```

10      1  MDGRTQTLQA  QTLGLISAAA  IILILILIVK  FRIRALLTLV  IASLLTALAT
      51  GLPTGSIVND  VLVKNFGGTL  GGVALLVGLG  AMLGRVETS  GGAQSLADAL
     101  IRMFGEKRAP  FAPGVASLIF  GFPIFFDAGL  IVMLPIVFAT  ARRMKQDVLP
     151  FALASVGAFS  VMHVFLPPHP  GPIAASEFYG  ANIGQVLILG  LPTAFITWYF
     201  SGYMLGKVLG  RAIHVPVPEL  LSGGTQSDP  PKEPAKAGTV  VAVMLIPMLL
     15  251  IFLNTGVSAL  ISEKLVSAD  TWVQTAKMIG  STPVALLISV  LAALLVLGRK
     301  RGESESTLEK  TVDGLAPAC  SVLITGAGG  MFGGVLRASG  IGKALADSMA
     351  DLGIPVLLGC  FLVALALRIA  QGSATVALTT  AAALMAPAVA  AAGFTDWQLA
     401  CIVLATAAGS  VGCSEFND  FWLVGRLLSDM  DVPTTLKTWT  VNQTLLAFIG
     451  FALSALLFAI  V*

```

20 Further work revealed a variant gonococcal DNA sequence <SEQ ID 591>:

```

      1  ATGGACGGCC  GGACACAGAC  GCTGTCCGCG  CAAACCTTGT  TGGGCATTTC
     51  GGCGGCGGCA  ATCATCCTCA  TTCTGATTTT  AATCGTCAAA  TTCCGCATCC
     101  GCGCGCTGCT  GACACTGGTC  ATCGCCAGCC  TGCTGACGGC  TTTGGCAACC
     151  GGTTTGCCCA  CAGGCAGCAT  CGTCAACGAC  GACTGGTCA  AAAACTTCGG
     201  CGGCACGCTC  GGCGGCGTGG  CGCTTCTGGT  CGGTCTGGGC  GCAATGCTCG
     25  251  GACGTTTGGT  AGAAACATCC  GGCGGCGCAC  AGTCGCTGGC  GGACGCGCTG
     301  ATCCGGATGT  TCGGCGAAAA  ACGCGCACCG  TTCGCTCCGG  GCGTTGCCTC
     351  GCTGATTTTC  GGCTTCCCGA  TTTCTTCGA  TGCCGGACTA  ATCGTCATGC
     401  TGCCCATCGT  ATTGCGCACC  GCACGGCGCA  TGAAACAGGA  CGTACTGCCC
     30  451  TTCGCGCTTG  CCTCCGTCGG  CGCATTTTCC  GTCATGCACG  TCTTCCTGCC
     501  GCGCCATCCG  GGCCCGATTG  CCGCTTCCGA  ATTTTACGGC  GCGAACATCG
     551  GCCAGTTTT  GATTTTGGGT  CTGCCGACCG  CCTTCATCAC  ATGGTATTTT
     601  AGCGGCTATA  TGCTCGGCAA  AGTGTGGGG  CGCGCCATCC  ATGTTCCCGT
     651  TCCGAACTG  CTCAGCGGCG  GCACGCAAGA  CAGCGACCCG  CCGAAAGAAC
     35  701  CTGCCAAAGC  AGGAACGTC  GTCGCCGTC  TGCTGATTCC  CATGCTGCTG
     751  ATTTTCTCTG  ATACCGCGGT  ATCAGCCCTC  ATCAGCGAAA  AACTCGTAAG
     801  TGCGGACGAA  ACTTGGGTTC  AGACGGCAAA  AATGATCGGT  TCGACACCTG
     851  TCGCCCTTCT  GATTTCCGTA  TTGGCCGCAC  TGTGGTCTT  GGGACGCAAA
     901  CGCGGCGAAA  GCGGCAGCAC  GTTGGAAAAA  ACCGTGGACG  GCGCACTCGC
     40  951  CCCCGCCTGT  TCCGTGATTC  TGATTACCGG  CGCGGGCGGT  ATGTTCCGGC
     1001  GCGTTTTCG  CGCTTCCGGC  ATCGGCAAGG  CACTCGCCGA  CAGCATGGCG
     1051  GATTTGGGCA  TTCCCGTCCT  TTTGGGCTGC  TTCTTGTCTG  CCTTGGCACT
     1101  GCGTATCGCG  CAAGGTTTCG  CAACCGTCGC  CCTGACCACA  GCCGCCGCGC
     1151  TGATGGCTCC  TGCGGTTGCC  GCCGCCGGCT  TTACCGACTG  GCAGCTCGCC
     45  1201  TGATATCGAT  TGGCAACGGC  GGCAGGTTTC  GTCGTTGCA  GCCACTCAA
     1251  CGACTCCGGC  TTCTGGCTGG  TCGGCCGCT  CTTGGATATG  GACGTACCGA
     1301  CCACGCTGAA  AACCTGGACG  GTCAACCAAA  CCCTCATCGC  ATTCATCGGC
     1351  TTTGCCTTGT  CCGCACTGCT  GTTGGCATC  GTCTGA

```

This corresponds to the amino acid sequence <SEQ ID 592; ORF140ng-1>:

```

50      1  MDGRTQTLQA  QTLGLISAAA  IILILILIVK  FRIRALLTLV  IASLLTALAT
     51  GLPTGSIVND  VLVKNFGGTL  GGVALLVGLG  AMLGRVETS  GGAQSLADAL
     101  IRMFGEKRAP  FAPGVASLIF  GFPIFFDAGL  IVMLPIVFAT  ARRMKQDVLP
     151  FALASVGAFS  VMHVFLPPHP  GPIAASEFYG  ANIGQVLILG  LPTAFITWYF
     201  SGYMLGKVLG  RAIHVPVPEL  LSGGTQSDP  PKEPAKAGTV  VAVMLIPMLL
     55  251  IFLNTGVSAL  ISEKLVSAD  TWVQTAKMIG  STPVALLISV  LAALLVLGRK
     301  RGESESTLEK  TVDGLAPAC  SVLITGAGG  MFGGVLRASG  IGKALADSMA
     351  DLGIPVLLGC  FLVALALRIA  QGSATVALTT  AAALMAPAVA  AAGFTDWQLA
     401  CIVLATAAGS  VGCSEFND  FWLVGRLLSDM  DVPTTLKTWT  VNQTLLAFIG
     451  FALSALLFAI  V*

```

60 ORF140ng-1 and ORF140-1 show 96.3% identity over 461aa overlap:

```

orfl140ng-1.pep MDGRTQTLQAQTLGLISAAAIIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND
                  ||| |||||:|||||:|||||:|||||:

```

-337-

```

orfl140-1      MDGWTQTLQAQTLGISAIAIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND
orfl140ng-1.pep VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF
5 orfl140-1      ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
orfl140ng-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLFFALASVGAFSVMHVFLPPHPGPIAASEFYG
10 orfl140-1      GFPIFFDAGLIVMLPIVFATARRMKQDVLFFALASIGAFSVMHVFLPPHPGPIAASEFYG
orfl140ng-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRAIHVPVPELLSGGTQDSDPPKEPAKAGTV
orfl140-1      ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVVPPELLSGGTQDNDLPKEPAKAGTV
15 orfl140ng-1.pep VAVMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKMIGSTPVALLISVLAALLVLGRK
orfl140-1      VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLAALLVLGRK
orfl140ng-1.pep RGESGSTLEKTVDGALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
20 orfl140-1      RGESGSALEKTVDGALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
orfl140ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEHNDG
25 orfl140-1      FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEHNDG
orfl140ng-1.pep FWLVGRLLDMDVPTTLKTWTVNQTIAFIGFALSALLFAIV
orfl140-1      FWLVGRLLDMDVPTTLKTWTVNQTIALIGFALSALLFAIV

```

30 Furthermore, ORF140ng-1 is homologous to an *E.coli* protein:

```

gi|882633 (U29579) ORF_o454 [Escherichia coli] >gi|1789097 (AE000358) o454;
This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
protein GNTP_BACLI SW: P46832 [Escherichia coli] Length = 454
Score = 210 bits (529), Expect = 1e-53
35 Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)

Query: 88 ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147
E SGGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K
40 Sbjct: 80 EHSGGAESLANYFSRKLGDKRTIAALTAAFFLGIPVFFDVGFIILAPIIYGFQAKAKIS 139

Query: 148 VLPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLILGLPTAFITWYFSGYMLGK 207
L F L G +HV +PPHPGP+AA+ A+IG + I+G+ + I GY K
Sbjct: 140 PLKFGLPVAGIMLTVHVAVPPHPGPVAAAGLLHADIGWLTIIIGIAIS-IPVGVVGYFAAK 198

45 Query: 208 VLGRAIHVPVPELL-----SGGTQDSDPPKEPAKAGTVVAVMLIPMLLIFLNTGV 257
++ + + E+L G T+ SD P A V ++++IP+ +I T
Sbjct: 199 IINKRQYAMSVEVLEQMQLAPASEEGATKLSDKINPPGVA-LVTSILIVPIAIIMAGT-- 255

50 Query: 258 SALISEKLVSADETWVQTAKMIGSTPXXXXXXXXXXXXXXXXXGRKRGESGSTLEKTVDGALA 317
+S L+ + T ++IGS +RG S + AL
Sbjct: 256 ---VSATLMPPSHPLLGLTLQLIGSPMVALMIALVLAFWLLALRRGWSLQHTSDIMGSALP 312

Query: 318 PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFLVALALRIAQGSXXXX 377
A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L F+++LALR +QGS
55 Sbjct: 313 TAAVVILVTGAGGVFGKVLVESGVGKALANMLQMIDLPPLPAFIISLALRASQGS--AT 370

Query: 378 XXXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSEHNDSGFWLVGRLLDMDVPTTLK 437
G Q + LA G +G SH NDSGFW+V + L + V LK
60 Sbjct: 371 VAILTTGGLLSEAVMGLNPIQCVLVTLAACFGGLGASHINDSGFWIVTKYLGSLVADGLK 430

Query: 438 TWTVNQTIAFIGFALSALLFAIV 461
TWTV T++ F GF ++ ++A++
Sbjct: 431 TWTVLTTILGFTGFLITWCVWAVI 454

```

Based on this analysis, including the identification of the presence of a putative leader sequence

65 (double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

5      1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCAAACA
      51  TTTGCTGTGC CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
     101  GCGTATTTT TGCCTTATC GGAAGTACTT CCTGCGGCTT TGCCGGTTTC
     151  AACTTTTGG GCAGACACCA CGGGCGCAC. GTCGTCCTGA TTCTCATCGG
     201  CTGTATCGGG CTGATTCCAG TTGCCCATTT CCTCAACCCC GCTGCCGCCG
     251  CCTTTGCCGC CGCCGGACTG GTGCTGCACG GTTATCTTTT GGCTCGCCGG
     301  CGCGTGATTG CCGCCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
     351  GTTGGCAGCA GCTTATCCGG CAGCATTTGC CCTGATGCTG CCCTTGCCCG
     401  TACTGATGTT TTTCCGTCGG ..

```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```

15      1  ..DFGISPVYLW VAAAFKHLLS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
      51  NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAFAAAGL VLGYSYLARR
     101  RVIAASFLLG TGWTLMSLAA AYPAAAFALML PLPVLMEFRP ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```

20      1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CGGCCCGCCA AAACCCACGA
      51  AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
     101  TGTTTTCCCA CGATTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
     151  GTCGAAGCAC TGGCAGGCAG CCCCACCCCG TTGGTTGCCG ATCTGTTCCG
     201  TCAAACCGAT TTCGCGATAC CGCCCGTGTA TCTTTGGGTT GCCGCCGCGT
     251  TCAAACATTT GCTGTCGCCG TGGGCTGCCG ACTCATACGA TGCCGCACGC
     301  TTGTCAGGCG TATTTTTTGC CGTTATCGGA CTGACTTCCT GCGGCTTTGC
     351  CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCTGATTC
     401  TCATCGGCTG TATCGGGCTG ATTCAGTTG CCCATTTCCT CAACCCCGCT
     451  GCCGCCGCCT TTGCCGCCG CGGACTGGTG CTGCACGGTT ATTCTTTGGC
     501  TCGCCGGGCG GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGCTGGACGC
     551  TGATGTCGTT GGCAGCAGCT TATCCGGCAG CATTTGCCCT GATGCTGCCG
     601  TTGCCCGTAC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTGATGTT
     651  GACGGCAGTC GCCTCACTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
     701  CGCTGCTCTT GGCAAAAACG CAGCCCGCGC TGTTCCGCGA ATGGCTCGAC
     751  TATCACGTTT TCGGTACGTT CGGCGCGGTG CGGCACGTT AGACGGCATT
     801  CAGTTTGTGT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCCGCGC
     851  TGCCGCTGCG GGTTTGGACG GTTTGCCGCA CGCGCCTGTT TTCGACCGAC
     901  TGGGGGATTT TGGCGTCTGT CTGGATGCTT GCCGTTTGG TGCTGCTTGC
     951  CGTCAATCCG CAGCGTTTTT AGGATAACCT CGTCTGGCTG CTTCGCCCGC
    1001  TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGCGG CGGCGCGGCG
    1051  GCGTTTGTCA ACTGTTCCGG CATTATGGCG TTCGGACTGT TTGCCGTGTT
    1101  CCTGTGGACG GGCTTTTTTC CCATGAATTA CGGCTGGCCC GCCAAGCTTG
    1151  CGCAACGCGC CGCCTATTTT AGCCCGTATT ATGTTCTCTG TATCGATCCC
    1201  ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTGGC TGTGGGCGAT
    1251  TACCCGGAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
    1301  GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
    1351  GACGCGGCGA AAAGCCACGC GCCGTCGTC CGGAGTATGG AGGCATCGCT
    1401  TTCCCCGGA TTTAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
    1451  TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGCACA
    1501  TTGCCGCACC GCGTCGGCGA TGTACAATGC CGCTACCGCA TCGTCCTCTT
    1551  GCCCAAAAT GCGGATGCGC CGCAAGGCTG GCAGACGGTT TGGCAGGGTG
    1601  CGCGTCCGCG CAACAAAGAC AGTAAGTTCT CACTGATACG GAAAATCGGG
    1651  GAAAATATAT AA

```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

```

55      1  MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
      51  VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAFAKHLSP WAADSYDAAR
     101  FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLLILIGCIG LIPVAHFLNPA
     151  AAAFAAAGLV LHGYSYLARR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
     201  LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD

```

-339-

251 YHVFGTFGGV RHVQTAFSLF YYLKNLLWFA LPALPLAVWT VCRTRLFSTD  
 301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDLSLRGAA  
 351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFE SPYYVPDIDP  
 401 IPMAVAVLFT PLWLWATRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL  
 451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIGIGGGDLH TRIVWTQYGT  
 501 LPHRVGDVQC RYRIVLLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG  
 551 ENI\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N. meningitidis*:

```

                                10      20      30
orfl41.pep                      DFGISPVYLWVAAAFKHLSPWAADSYDVA
15 orfl41a      WNPDEPAVYTAVEALAGSPTPLVAHLFGQIDFGIPPVYLWVAAAFKHLSPWAADPYDAA
                   40      50      60      70      80      90

                                40      50      60      70      80      90
20 orfl41.pep      RFAGVFFAVIGLTSCGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAFAAAGL
orfl41a      RFAGVFFAVVGLTSCGFAGFNFLGRHHGRSVVLILIGCIGLIPTVHFLNPAAAFAAAGL
                   100     110     120     130     140     150

                                100     110     120     130     140
25 orfl41.pep      VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRP
orfl41a      VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRMLMLTA
                   160     170     180     190     200     210

30 orfl41a      VASLAFALPLMTVYPLLLAKTQPALFAQWLDDHVFGTFGGVRHIQTAFSLFYFLKNLLWF
                   220     230     240     250     260     270

```

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

```

1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
35 51  AAAGCCGTGG CTGTTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
101 TGTTTTCCCA CGATTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
151 GTCGAAGCAC TGGCAGGCAG CCCACCCCT TTGGTTGCC ATCTGTTCCG
201 TCAAATCGAT TTCGGCATA CGCCCGTGTA TCTTTGGGTT GCCGCCGCGT
251 TCAAACATTT GCTGTCGCCG TGGGCTGCCG ACCCGTATGA TGCCGCACGC
40 301 TTGCGCGCG TGTTTTTCGC CGTTGTGCGA CTGACTTCCT GCGGCTTTGC
351 CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATT
401 TCATCGGCTG TATCGGCTG ATTCCGACCG TACACTTCT CAACCCCGCT
451 GCGCGCGCCT TTGCGCGCG CGGACTGGTG CTGCACGGTT ATTCTTTGGC
501 TCGCCGGCGC GTGATTGCCG CCTCTTTCT GCTCGGTACG GGTGGACGC
551 TGATGTCGTT GGCAGCAGCT TATCCGGCG CATTTGCCCT GATGCTGCCC
45 601 CTGCCCGTGC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
651 GACGGCAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
701 CGTGCTCTT GGCAAAAACG CAGCCCGCGC TGTCGCGCA ATGGCTCGAC
751 GATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCACATTC AGACGGCATT
801 CAGTTTGTGT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCTGCGC
50 851 TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CGCGCTGTT TTCGACCGAC
901 TGGGGGATTT TGGGCGTCTG CTGGATGCTT GCCGTTTGG TGCTGCTTGC
951 CGTCAATCCG CAGCGTTTTT AGGATAACCT CGTCTGGCTG CTTCCGCGCG
1001 TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGACG CGGCGCGGCG
1051 GCGTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGACTGT TTGCCGTGTT
55 1101 CCTGTGGACG GGCTTTTTTC CCATGAATTA CGGCTGGCCC GCCAAGCTTG
1151 CCGAACGCGC CGCCTATTTT AGCCCGTATT ATGTTCTCTGA TATCGATCCC
1201 ATCCGATGG CGGTTGCCGT ACTGTTTACA CCCTGTGGC TGTGGGCGAT
1251 TACCCGCAAA AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
1301 GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
60 1351 GACGCGGCGA AAAGCCACGC GCCCGTCGTC CGGAGTATGG AGGCATCGCT
1401 TTCCCGGAA TTAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGACA
1451 TAGGCGGCGG CGACCTACAC ACGCGGATG TTTGGACGCA GTACGGCACA
1501 TTGCCGCACC GCGTCGGCGA TGTACAATGC CGCTACCGCA TCGTCCGCTT
1551 GCCCCAAAAC GCGGATGCGC CGCAAGGCTG GCAGACGGTC TGGCAGGGTG

```

1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG  
 1651 GAAAATATAT TAAAAACAAC AGATTGA

This encodes a protein having amino acid sequence <SEQ ID 598>:

1 MLTYTPPDAR PPAKTHEKWP LLLMAFAWL WPGVFSHDLW NPDEPAVYTA  
 5 51 VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAFAKHLSP WAADPYDAAR  
 101 FAGVFFAVVG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPTVHFLNPA  
 151 AAFAAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP  
 201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD  
 251 DHVFGTFGGV RHIQTAFSLF YYLKNLLWFA LPALPLAVWT VCRTLRFSTD  
 10 301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDLRLRGAA  
 351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYYVPDIDP  
 401 IPMAVAVLFT PLWLWAI TRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL  
 451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIDIGGGDLH TRIVWTQYGT  
 501 LPHRVGDVQC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKGTG  
 15 551 ENILKTTD\*

ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

orf141a.pep MLTYTPPDARPPAKTHEKWP LLLMAFAWLWPGVFSHDLWNPDEPAVYTA VEALAGSPTP  
 20 orf141-1 MLTYTPPDARPPAKTHEKWP LLLMAFAWLWPGVFSHDLWNPDEPAVYTA VEALAGSPTP  
 orf141a.pep LVAHLFGQIDFGIPPVYLWVAAFAKHLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN  
 orf141-1 LVAHLFGQIDFGIPPVYLWVAAFAKHLSPWAADSYDAARFAGVFFAVVGLTSCGFAGFN  
 25 orf141a.pep FLGRHHGRSVVLILIGCIGLIPTVHFLNPA AAFAAGLV LHGYSLARRR VIAASFLLGT  
 orf141-1 FLGRHHGRSVVLILIGCIGLIPTVHFLNPA AAFAAGLV LHGYSLARRR VIAASFLLGT  
 30 orf141a.pep GWTLMSLAAA YPAAFALMLP LPVLMFFRPW QSRRLMLTAVASLAFALPLMTVYPLLLAKT  
 orf141-1 GWTLMSLAAA YPAAFALMLP LPVLMFFRPW QSRRLMLTAVASLAFALPLMTVYPLLLAKT  
 orf141a.pep QPALFAQWLD DHVFGTFGGVRHIQTAFSLFYYLKNLLWFALPALPLAVWTVCRTLRFSTD  
 35 orf141-1 QPALFAQWLD DHVFGTFGGVRHVQTAFSLFYYLKNLLWFALPALPLAVWTVCRTLRFSTD  
 orf141a.pep WGILGVVWMLAVLVLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRLRGAAAFVNWFGIMA  
 40 orf141-1 WGILGVVWMLAVLVLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRLRGAAAFVNWFGIMA  
 orf141a.pep FGLFAVFLWTGFFAMNYGWP AKLAERAAYFSPYYVPDIDP IPMAVAVLFTPLWLWAI TRK  
 orf141-1 FGLFAVFLWTGFFAMNYGWP AKLAERAAYFSPYYVPDIDP IPMAVAVLFTPLWLWAI TRK  
 45 orf141a.pep NIRGRQAVTNWAAGVTLT WALLMTLFLPWL DAAKSHAPVRSMEASLSPE LKRELSDGIE  
 orf141-1 NIRGRQAVTNWAAGVTLT WALLMTLFLPWL DAAKSHAPVRSMEASLSPE LKRELSDGIE  
 orf141a.pep CIDIGGGDLHTRIVWTQYGT LPHRVGDVQC RYRIVRLPQNADAPQGWQTVWQGARPRNKD  
 50 orf141-1 CIDIGGGDLHTRIVWTQYGT LPHRVGDVQC RYRIVRLPQNADAPQGWQTVWQGARPRNKD  
 orf141a.pep SKFALIRKGTGENI  
 55 orf141-1 SKFALIRKGTGENI

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

60 orf141.pep DFGISPVYLWVAAFAKHLSPWAADSYDVA 30  
 orf141ng WNPAAEPVYTA VEALAGSPTPLVAHLFGQIDFGIPPVYLWVAAFAKHLSPWAAHPYDAA 126



-341-

	orf141.pep	RFAGVFFAVIGLTS CGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL	90
	orf141.ng	RFAGVFFAVIGLTS CGFAGFNFLGRHHGRSVVLIHIGCIGLIPVAHFFNPAAAAFAAAGL	186
5	orf141.pep	VLHGYSLARRRVIAASFLLTGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRP	140
	orf141.ng	VLHGYSLARRRVIAASFLLTGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRPQSRRLMLTA	246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

10	1	MPSEAVSARP	LCEYLLHLAI	RPFLTLMLT	YTPPDARPPA	KTHEKPWLLL
	51	LMAFAWLWPG	VFSHDLWNPA	EPAVYTAVEA	LAGSPTPLVA	HLFGQTDFGI
	101	PPVYLWVAAA	FKHLLSPWAA	HPYDAARFAG	VFFAVIGLTS	CGFAGFNFLG
	151	RHHGRSVVLI	HIGCIGLIPV	AHFFNPAAAA	FAAAGLVLHG	YSLARRRVIA
	201	ASFLLTGTGWT	LMSLAAAYPA	AFALMLPLPV	LMFFRPWQSR	RLMLTAVASL
15	251	AFALPLMTVY	PLLLAKTQPA	LFAQWLNYHV	FGTFGGVRHI	QRAFSLFHYL
	301	KNLLWFAPPG	LPLAVWTVCR	TRLFSTDWGI	LGIVWMLAVL	VLLAFNPQRF
	351	QDNLVWLLPP	LALFGAAQLD	SLRRGAAAFV	NWFGIMAFGL	FAVFLWTGFF
	401	AMNYGWPAKL	AERAAYFSPY	YVPDIDPIPM	AVAVLETPLV	LWAITRKNIR
	451	GRQAVTNWAA	GVTTLTWALL	TLFLPWLDAA	KSHAPVVRSM	EASFSPCLKR
20	501	ELSDGIECIG	IGGGDLHTRI	VWTQYGTLPH	RVGDVRCRYR	IVRLPQNADA
	551	PQGWQTVWQG	ARPRNKDSKF	ALIRKIGENI	LKTTD*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

	1	ATGCTGACCT	ATACCCCGCC	CGATGCCCGC	CCGCCCGCCA	AAACCCACGA
	51	AAAACCGTGG	CTGCTGCTGT	TGATGGCGTT	TGCTGGCTG	TGGCCCGGCG
25	101	TGTTTCCCA	CGATTGTGG	AATCCTGCCG	AACCTGCCGT	CTATACCGCC
	151	GTCGAAGCAC	TGGCAGGCAG	CCCCACCCCG	TTGGTTGCCG	ATCTGTTCGG
	201	TCAAACCGAT	TTCGGCATA	CGCCCGTGTA	TCTTTGGGTT	GCCGCCGCAT
	251	TCAAACATTT	GCTGTCGCCG	TGGGCAGCCG	ACCCGTATGA	TGCCGCACGC
	301	TTTGCAGGCG	TATTTTTTGC	CGTTATCGGA	CTGACTTCTT	GCGGCTTTGC
30	351	CGGTTTCAAC	TTTTTTGGCA	GACACCACGG	GCGCAGCGTT	GTTTTAATCC
	401	ATATCGGCTG	TATCGGCTG	ATTCCGGTTG	CCCATTTTCT	CAATCCcgcc
	451	gcccgcgcct	tTGCCGCCCG	CGGACTGGTG	CTGCacgget	actcgctgCG
	501	AGCCCGGCGC	GTGATTgccc	cctctTtccT	GCTCGGTACG	GTTTGACCGT
	551	TGATGTCGCT	GGCGGCAGCT	TATCCGGCGG	CGTTTGCCTG	GATGCTGCCC
35	601	CTGCCCGTGC	TGATGTTTTT	CCGTCCGTGG	CAAAGCAGGC	GTTTGATGTT
	651	GACGGCAGTC	GCCTCGCTTG	CCTTTGCCCT	GCCGCTTATG	ACCGTTTACC
	701	CGCTGCTctt	gGCAAAAACG	CAGCCCGCGC	TGTTTGC CGA	ATGGCTCAAC
	751	TATCACGTTT	TCCGTACGTT	cggcgGCGTG	CGGCACaTTC	AGAggGCatT
	801	Cagtttgttt	cactatctgA	AAaatctgct	ttggttcgca	ccgcccgggC
40	851	TGCCCGTGGC	GGTTTGGACG	GTTTGCCGCA	CACGCCTGTT	TTCGACCGAC
	901	TGGGGGATTT	TGGGCATTGT	CTGGATGCTT	GCCGTTTTTG	TGCTGCTCGC
	951	CTTTAATCCG	CAGCGTTTTT	AAGACAACCT	CGTCTGGCTG	CTGCCGCCGC
	1001	TTGCCCTGTT	CGGCGCGGCG	CAACTGGACA	GCCTGAGGCG	CGGCGCGGCG
	1051	GCTTTTGTCA	ACTGGTTCGG	CATTATGGCG	TTCCGGCTGT	TTGCCGTGTT
45	1101	CCTGTGGACG	GGCTTTTTTC	CCATGAATTA	CGGCTGGCCC	GCCAAGCTTG
	1151	CCGAACGCGC	CGCCTACTTC	AGCCCGTATT	ACGTTCCCGA	CATCGATCCC
	1201	ATTCGATGCG	CGGTTGCCGT	ACTGTTTACA	CCCTTGTGGC	TGTGGGCGAT
	1251	TACCCGGAAG	AACATACGCG	GCAGGCAGGC	GGTTACCAAC	TGGGCGGCGG
50	1301	GCGTTACCC	GACCTGGGCT	TTGCTGATGA	CGCTGTTTCT	GCCGTGGCTG
	1351	GACGCGGCGA	AAAGCCACGC	GCCCGTCGTC	CGGAGTATGG	AGGCATCGTT
	1401	TTCCCGGAA	TTAAACGGG	AGCTTTCAGA	CGGCATCGAG	TGTATCGGCA
	1451	TAGGCGGCGG	CGACCTGCAC	ACGCGGATTG	TTTGGACGCA	GTACGGCACA
	1501	TTGCCGCACC	GCGTCGCGCA	TGTCGTTTGC	CGCTACCGTA	TGCTCCGCTT
	1551	GCCCCAAAAC	GCGGATGCGC	CGCAAGGCTG	GCAGACGGTC	TGGCAGGGTG
55	1601	CGGCGCGGCG	CAACAAAGAC	AGTAAGTTTG	CACTGATACG	GAAAATCGGG
	1651	GAAAATATAT	TAAAAACAAC	AGATTGA		

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

	1	MLTYTPPDAR	PEAKTHEKPW	LLLLMAFAWL	WPGVFSHDLW	NPAEPAVYTA
	51	VEALAGSPTP	LVAHLFGQTD	FGIPPVYLWV	AAAFKHLLSP	WAADPYDAAR
60	101	FAGVFFAVIG	LTSCGFAGFN	FLGRHHGRSV	VLIHIGCIGL	IPVAHFLNPA
	151	AAAFAAAGLV	LHGYSLARRR	VIAASFLLTGT	GWTLMSLAAA	YPAAFALMLP
	201	LPVLMEFRPW	QSRRLMLTAV	ASLAFALPLM	TVYPLLLAKT	QPALFAQWLN
	251	YHVFGTFFGV	RHIQRAFSLF	HYLKNLLWFA	PPGLPLAVWT	VCRTRLFSTD
	301	WGILGIVWML	AVLVLLAFNP	QRFQDNLVWL	LPPLALFGAA	QLDSLRRGAA

-342-

5 351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP  
 401 IPMAVAVLFT PLWLWAI TRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL  
 451 DAAKSHAPVV RSMEASF SPE LKRELS DGIE CIGIGGGDLH TRIVWTQYGT  
 501 LPHRVGDVRC RYRIVRLPQN ADA PQGWQTV WQGARPRNKD SKFALIRKIG  
 551 ENILKTTD\*

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

10 orf141ng-1.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAEPAVYTAVEALAGSPTP  
 orf141-1 MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNDEPAVYTAVEALAGSPTP  
 orf141ng-1.pep LVAHLFGQTDGFI PPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVIGLTS CGFAGFN  
 orf141-1 LVAHLFGQTDGFI PPVYLWVAAAFKHLLSPWAADSYDAARFAGVFFAVIGLTS CGFAGFN  
 15 orf141ng-1.pep FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLV LHGYS LARRRVIAAS FLLGT  
 orf141-1 FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAAFAAAGLV LHGYS LARRRVIAAS FLLGT  
 20 orf141ng-1.pep GWTLM SLAAAYPAA FALMLPLPVLMEFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT  
 orf141-1 GWTLM SLAAAYPAA FALMLPLPVLMEFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT  
 25 orf141ng-1.pep QPALFAQWLNHYHVF GTFGGVRHIQRAFS LFHYLKNLLWFAPPGLPLAVWTVCTR LFTSD  
 orf141-1 QPALFAQWLDYHVF GTFGGVRHVQTAFSLFY LKNLLWFALPALPLAVWTVCTR LFTSD  
 30 orf141ng-1.pep WGILGIVWMLAVLVLLAFNPQR FQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 orf141-1 WGILGVVWMLAVLVLLAVNPQR FQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA  
 orf141ng-1.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPM AVAVLFTPLWLWAI TRK  
 orf141-1 FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPM AVAVLFTPLWLWAI TRK  
 35 orf141ng-1.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSM EASF SPELKRELS DGIE  
 orf141-1 NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAAKSHAPVVRSM EASF SPELKRELS DGIE  
 40 orf141ng-1.pep CIGIGGGDLHTRIVWTQYGTLP HRVGDVRCRYRIVRLPQNADAPQGWQTVWQ GARPRNKD  
 orf141-1 CIGIGGGDLHTRIVWTQYGTLP HRVGDVQCRYRIVLLPQNADAPQGWQTVWQ GARPRNKD  
 orf141ng-1.pep SKFALIRKIGENILKTTDX  
 45 orf141-1 SKFALIRKIGENIX

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## Example 72

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

1 ..CAATCCGCCA AATGGTTATC GGGCCAAACT CTAGTCGCCA CAGCAATTGG  
 51 GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA  
 101 CCGGCCGCGC ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA  
 151 AGCGGTTTTTC AGGTAGGCTA TACGTTTTAA

55 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

1 ..QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA  
 51 SGFQVGYTF\*

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

-343-

1 ATGGATAATT CGGGTAGTGA GCGCACAGGA AAATACCAAG GAAATATCAC  
 51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTT TATGTAATTT  
 101 ATGGACGTTT GATTGGCGGT ACGCCCGATG AGGAAAGTTT TGACGGCCAT  
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT  
 201 CGGTAAATGG ACATGGGCGT TCAATCACAA TGGCTACCGT TACCATCAGG  
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAT  
 301 ACTGATTTTC GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC  
 351 CTATCTCGGT GTAAACTGT GGATGAGGGA AACAAAAAGT TACATTGATG  
 401 ATGCCGAAC T GACTGTACAA CGGCGTAAAA CTGCGGGTTG GTTGGCAGAA  
 451 CTTTCCCACA AAGAATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA  
 501 ATATAAACGC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG  
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT  
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTTGCCT ATGACACATC  
 651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAAACTGG  
 701 CTATCGGCGG ACACACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG  
 751 TCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTTAA  
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGTAGGACAT GTTTCAGGAC  
 851 AATCCGCCAA ATGTTTATCG GGCCAAACTC TAGTCGGCAC AGCAATTGGG  
 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC  
 951 CGGCCGCGCA TTGAAAAAGC CCGAATTTTT CCAATCAAGG AAATGGGCAA  
 1001 GCGGTTTTCA GGTAGGCTAT ACGTTTTTAA

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

1 MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH  
 51 RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQAUSGLSE VYDYNKSYN  
 101 TDFGFNRLLY RDAKRKTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE  
 151 LSHKEYIGRS TADFKLKYYR GTGMKDALRA PEEAFEGETS RMKIWTASAD  
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL  
 251 SAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLVGTAG  
 301 IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

orf142.pep QSAKWLSGQTLVGTAGTIGRQIKLGGNLHY 30  
 orf142ng RGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLGTAIGRQIKLGGNLHY 313  
 orf142.pep DIFTGRALKKPEFFQSRKWASGFQVGYTF 59  
 orf142ng DIFTGRALKKPEYFQTKKVVTFQVGYSF 342

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

1 ATGGATAATT CGGGTAGTGA GCGCACAGGA AAATACCAAG GAAATATCAC  
 51 TTTCTCTGCC GACAATCCTT TTGGACTGAG TGATATGTTT TATGTAATTT  
 101 ATGGACGTTT AATTGGCGGT ACGCCCGATG AGGAAATTTT TGACGGCCAT  
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT  
 201 CGGTAAATGG ACATGGGCGT TCAATCACAA TGGCTACCGT TACCATCAGG  
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAC  
 301 ACTGATTTTC GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC  
 351 CTATCTCAGT GTAAACTGT GGACGAGGGA AACAAAAAGT TACATTGATG  
 401 ATGCCGAAC T GACTGTACAA CGGCGTAAAA CCACAGGTTG GTTGGCAGAA  
 451 CTTTCCCACA AAGGATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA  
 501 ATATAAACAC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG  
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT  
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTTGCCT ATGACACATC  
 651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAAACTGG  
 701 CTATCGGCGG ACACACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG  
 751 CCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTTAA  
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGTAGGACAT GTTTCAGGAC  
 851 AATCCGCCAA ATGTTTATCG GGCCAAACTC TAGCCGGCAC AGCAATTGGG  
 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC  
 951 CGGCCGTGCA TTGAAAAAGC CCGAATATTT TCAGACGAAG AAATGGGTAA

1001 CGGGGTTTCA GGTGGGTAT TCGTTTGA

This encodes a protein having amino acid sequence <SEQ ID 608>:

1 MDNSGSEATG KYQGNITFSA DNPFGLSDMF YVNYGRSIGG TPDEENFDGH  
 51 RKEGGSNNYA VHYSAPFGKW TWA FNHNGYR YHQA VSGLSE VYDYNKSYN  
 101 TDFGFNRLLY RDAKRKTYLS VKLWTRETKS YIDDAELTVQ RRKTTGWLAE  
 151 LSHKGYIGRS TADFKLKYKH GTGMKDALRA PEEAFGEGTS RMKIWTASAD  
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL  
 251 PAERGWWYRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLAGTAIG  
 301 IRGQIKLGGN LHYDIFTGRA LKKPEYFQTK KWTGFGVGY SF\*

10 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

orfl42-1.pep MDNSGSEATGKYQGNITFSADNPLGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA  
 15 orfl42ng-1 MDNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA  
 orfl42-1.pep VHYSAPFGKWTWAFNHNHNGYRYHQA VSGLSEVYDYNKSYNTDFGFNRLLYRDAKRKTYLG  
 20 orfl42ng-1 VHYSAPFGKWTWAFNHNHNGYRYHQA VSGLSEVYDYNKSYNTDFGFNRLLYRDAKRKTYLS  
 orfl42-1.pep VKLWMRETKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLKYKRGTMKDALRA  
 orfl42ng-1 VKLWTRETKSYIDDAELTVQRRKTTGWLAE LSHKGYIGRSTADFKLKYKHGTGMKDALRA  
 25 orfl42-1.pep PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT  
 orfl42ng-1 PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT  
 orfl42-1.pep VRGFDGEMSLSAERGWWYRN DLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLVGTAIG  
 30 orfl42ng-1 VRGFDGEMSLPAERGWWYRN DLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIG  
 orfl42-1.pep IRGQIKLGGNLHYDIFTGRALKKPEFFQSRKWASGFGVGYTF  
 35 orfl42ng-1 IRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGFGVGYSF

In addition, ORF142ng is homologous to the HecB protein of *E.chrysanthemi*:

gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558  
 Score = 119 bits (295), Expect = 3e-26  
 Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)  
 40 Query: 2 DNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61  
 DNSG ++TG+ Q N + + DN FGL+D +++ G S + + D + G  
 Sbjct: 230 DNSGQKSTGEEQLNGSLALDNVFGGLADQWFLSAGHS---SRFATSHDAESLQAG----- 280  
 45 Query: 62 HYSAPFGKWTWAFNHNHNGYRYHQA VSGLSEVYDYNKSYNTDFGFNRLLYRDAKRKTYLSV 121  
 +S P+G W +N++ RY + G S F +R+++RD KT ++  
 Sbjct: 281 -FSMPYGYWNLYGYNYSQSRYNRTFINRDFPWHSTGDSDFRSLRVVFRDGTMTAIAAG 339  
 50 Query: 122 KLWTRETKSYIDDAELTVQRRKTGWLAE LSHKGYIGRSTADFKLKYKHGTGMKDALRAP 181  
 R +Y++ + L RK + ++H + A F Y G +  
 Sbjct: 340 TFSQRTGNNYLNGLSLLPSSSRKLSSVSLGVNHSQKLWGGLATFNPTYNRVRWLGETDT 399  
 Query: 182 EEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHTV 241  
 +++ E + WT SA P Y S++ Q++ L ++L +GG ++  
 55 Sbjct: 400 DKSADEPRAEFNKWTLASYYHPV---TDSITYLGSLYGQYSARALYGSEQLTLGGESSI 456  
 Query: 242 RGFDGEMSLPAERGWWYRN DLSWQFKP----GHQLYLGA-DVGHVSGQSAKWLSGQTLAG 296  
 RGF E RG YWRN+L+WQ G+ ++ A D GH+ + +L G  
 60 Sbjct: 457 RGF-REQYTSGNRGAYWRNELNWQAWQLPVLGNVTFMAAVDGGHLYNHHKQDNSTAASLWG 515  
 Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGFGVGYSF 342  
 A+G+ + L + G + P + Q V G++VG SF

### 5 Example 73

10

1	ATGCGGACGA	AATGGTCAGC	AGTGAGAAGC	TGCTTACTTG	GgCGGACACC
51	GCCGACATCG	ATACCGCTTT	GAACCTGTTG	TACCGTTTGC	AAAACTTCGA
101	ATTCTCTAT	GGCGATGAA	ACGGTCATT	AGCGGCATC	AATTTGwCGG
151	ACGAGCAATT	CGCGTTGCTG	ATGGAACAAT	TGTCGGCGAG	CGGTAAAGCG
201	TTATTGGTCG	ATCGGAACGG	TCTGTATCTT	GCCACAGCCA	ATTTCCATCA
251	TGAGGCGCGG	GAAGAGTTTG	GGTTGTTGGC	GGCAGAAGTC	GCACAGATGG
301	AAAAGAAATA	CCGGCTGCTG	ATTAAGAACA	GC...	

```

15      1  MRTKWSAVRS  CTWADTADID  TALNLLYRLQ  KLEFLYGDEN  GHSDGINLXD
      51  EQLPLLMEQL  SGSGKALLVD  RNGLYLANAN  FHHEAAEELG  LLAAEVAQME
     101  KKYRLLIKNN  ..

```

20	1	ATGGAATCAA	CACCTTTCAC	ACAAGCAAAT	TTATATCCCC	GCCTGACTCC
	51	TGCCGGTGCA	TTTTATGCCG	TATCCAGCGA	TGCCCCCAGT	GCCGGTAAAA
	101	CTTTGTTGCA	TTCCTGTGTT	AAAGCAGATG	CGGACGAAAT	GGTCAGCAGT
	151	GAGAAGCTGC	TACTTGGGC	GGACACGCC	GACATCGATA	CCGCTTTGAA
25	201	CCTGTTGTAC	CGTTTGCAAA	AACTCGAATT	CCCTCATGGC	GATGAAACAC
	251	GTCATTGAGA	CGGCATCAAT	TTGTCGGACG	AGCAATTGCC	GTTGCTGATG
	301	GAACAATTGT	CCGGCAGCGG	TAAGGCGTTA	TTGGTCGATC	GGAACGGTCT
	351	GTATCTTGCC	AACGCCAATT	TCCATCATGA	GCGCGCGGAA	CAGTTGGGGT
30	401	TGTTGGCGGC	AGAAGTCGCA	CAGATGGAAA	AGAAATACCG	GCTGCTGATT
	451	AAGAACAAAC	TGTATATCAA	CAATAACGCT	TGGGGCGTTT	CGCATCCTTC
	501	CGGTGAGAGC	GAATTGACAT	TTTTCCTCAT	GTATATCGGT	TCAACCAAAT
	551	TTATTTTGGT	TATCGGCGGC	ATTCCCGATT	TGGGCAAAGT	GGCATTTGTT
	601	ACTTTGGGTAA	GGATTTTATA	CCGCCGTTAC	AGCAACCGCG	TGTAA

35

1	MESTLSLQAN	LYPRLTPAGA	FYAVSSDAPS	AGKTLHSL	KADADEMVSS
51	EKLLTWADTA	DIDTALNLLY	RLQKLEFLYG	DENGHSDDGIN	LSDEQLPLLM
101	EQLSGSGKAL	LVDRNGLYLA	NANFHHAEAE	ELGLLAAEVA	QMEKKYRLLI
151	KNNLYINNNNA	WGVCDSGQS	ELTFFFLYIG	STKFILVIGG	IPDLGKEAFV
201	TLVRILYRRY	SNRV*			

Homology with a predicted ORF from *N.meningitidis* (strain A)

```

                                     10      20      30
orfl43.pep                        MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFL
                                     | : : ||| ||||| ||||| |||||
45 orfl43a    GAFYAVSSDXPSAGKTLLHSLLKADADEMVSSEKLLTWAXTADIDTALNLLYRLQKLEFL
                20          30          40          50          60          70

                                     40      50      60      70      80      90
50 orfl43.pep    YGDENGHSDGINLXDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHHEEAEELGLLAAE
                    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

-346-

```

orfl43a      YGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHEAAEELGLLAAE
              80      90      100      110      120      130
5
orfl43.pep   100      110
              VAQMEKKYRLLIKNN
              |||||
orfl43a      VAQMEKKYRLXIKNNLYINNNAWGVCDPGQSELTFFFPLYIGSTKFILVIGGIPDLGKEA
              140      150      160      170      180      190

```

The complete length ORF143a nucleotide sequence <SEQ ID 613> is:

```

10      1  ATGGAATCAA  CANTTTCAC  ACAAGCAAAT  TTATATCNCC  GCCTGACTCC
      51  TGCCGGTGCA  TTTTATGCCG  TATCCAGCGA  TGNCCCCAGT  GCCGGTAAAA
     101  CTTGTGTGCA  CAGCCTGTTG  AAAGCGGATG  CGGACGAAAT  GGTNAGCAGT
     151  GAGAAGCTGC  TTACCTGGGC  GGANACCGCC  GACATCGATA  CCGCTTTGAA
     201  CCTGTTGTAC  CGTTTGCAAA  AACTCGAATT  CCTCTATGGC  GATGAAAACG
     251  GTCATTCAGA  CGGCATCAAT  TTGTCGGACG  AGCAATTGCC  GTTGCTGATG
     301  GAACAATTGT  CCGGCAGCGG  TAAGGCGTTA  TTGGTCGATC  GGAACGGTCT
     351  GTATCTTGCC  AACGCCAATT  TCCATCATGA  GGCGGCGGAA  GAGTTGGGGT
     401  TGTGGCGGC  AGAAGTCGCA  CAGATGGAAA  AGAAATACCG  GCTGCNNATT
     451  AAGAACAACC  TGTATATCAA  CAATAACGCT  TGGGGCGTTF  GCGATCCTTC
     501  CGGTCAGAGC  GAATTGACAT  TTTTCCCAT  GTATATCGGT  TCAACCAAAAT
     551  TTATTTTGGT  TATCGGCGGC  ATTCGCCATT  TGGGCAAAGA  GGCATTTGTT
     601  ACTTTGGTAA  GGATNTTATA  CCNCCNGTTA  CAGCAACCGC  GTGTAAAACT
     651  TGGGAGAGAG  GANGGGTTAT  GCAGCAATTA  TTGA

```

This encodes a protein having amino acid sequence <SEQ ID 614>:

```

25      1  MESTXSLQAN  LYXRLTPAGA  FYAVSSDXPS  AGKTLHSL  KADADEMVSS
      51  EKLLTWAXTA  DIDTALNLLY  RLQKLEFLYG  DENGHSDGIN  LSDEQLPLL
     101  EQLSGSGKAL  LVDRNGLYLA  NANFHHEAAE  ELGLLAAEVA  QMEKKYRLXI
     151  KNNLYINNNA  WGVCDPGQSQ  ELTFFPLYIG  STKFILVIGG  IPDLGKEAFV
     201  TLVRXLYXXL  QQPRVKLGRE  XGLCSNY*

```

ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

```

orfl43a.pep  MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTLHSLKADADEMVSSEKLLTWAXTA
orfl43-1     MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLHSLKADADEMVSSEKLLTWADTA
35
orfl43a.pep  DIDTALNLLYRLQKLEFLYGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
orfl43-1     DIDTALNLLYRLQKLEFLYGDENGHSDGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
40
orfl43a.pep  NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNNLYINNNAWGVCDPGQSELTFFFPLYIG
orfl43-1     NANFHHEAAEELGLLAAEVAQMEKKYRLLIKNNLYINNNAWGVCDPGQSELTFFFPLYIG
orfl43a.pep  STKFILVIGGIPDLGKEAFVTLVRXLY
45
orfl43-1     STKFILVIGGIPDLGKEAFVTLVRILY

```

#### Homology with a predicted ORF from *N. gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N. gonorrhoeae*:

```

50      orfl43.pep  MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYGDENGHSDGINLXDEQLPLLMEQL  60
      orfl43ng    MRTKWSAVRSCSRADTADIDTALNLLYRLQKLEFLYGDENGHSDGINLSDEQLPLLMEQL  60
55
      orfl43.pep  SGSGKALLVDRNGLYLANANFHEAAEELGLLAAEVAQMEKKYRLLIKNN  110
      orfl43ng    SGSGKALLVDRNGLYLANANFHESAEELGLLAAEVAQMEKKYRLIIRNNLYINNNAWGV  120

```

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

15

20

25

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

35

40

45

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

55

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLQRLQGL ADNKICAFW FVVRREFDEER VPQXAASMTF TTLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQTIVP XGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVX XQRPWM...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG
151 ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCGGTGTTTC ACCGTGTTTC
201 GGATTCGTTT GCATCTCTTC TCAACCAAAC CATGTGTCGG CAGGCGCGCG
10  251 ACATGGTGTG CACTATATAT AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCTGT ACCTCGCTGA TGCTGATTCG
351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCGG
15  451 CTGTCTTTGG GCGTGGGCGT TTCCTTTATG GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC GCGAGTGGTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAC CTTTCATGAC CTTTGTCTGT GGGGCTGTG CCGCTTCGTG
601 CCAAACGCTC TCGTTCCGCG GCGGCAGGCG TTTGTCGGGG CTTTGGCAAC
651 AGCGTTTTGT CTGGAACCGC GCGGCTCCCT CTCACCTTGG TATATGGGCA
20  701 ATTTGACGCG CTACCGCTCG ATTTACGGCG CGTTTGCCGC CGTGCCGTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
801 CCGTACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CAAAGCCTTG CCTGTTTCTG AGTTCAGACG
25  951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGTTGGGT GTTGAAAACG
1051 GGGCGGGATT CGATTGAGTT GAACGAATC TTCAAGCTCT TCGTTTACCG
1101 TCCGTTGCCT GTGGAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
30  1201 CAGGCGAAAA AACGGCAGTA G

```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLQRLQGL ADNKICAFW FVVRREFDEER VPQXAASMTF TTLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQTIVP QGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
35  151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFMT LLLWGLYRFV
201 PNRFPVPAQA FVGALATAFC LETARSLFTW YMGNFDGYRS IYGAFAPVFP
251 FLWLNLWLT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLARHGYYI SGROGWVLKT
40  351 GADSIELNEL FKLFFVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA
401 QAKKRQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N. meningitidis*:

```

45  orf144.pep  10      20      30      40      50      60
      MTFLQRLQGLADNKICAFW FVVRREFDEERVPQXAASMTFTTLLALVPVLTVMVAVASIF
      orf144a  10      20      30      40      50      60
      MTFLQRLQGLADNKICAFW FVVRREFDEERVPQXAASMTFTTLLALVPVLTVMVAVASIF

50  orf144.pep  70      80      90      100     110     120
      PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
      orf144a  70      80      90      100     110     120
      PVFDRWSDSFVSFVNQTIVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID

55  orf144.pep  130
      NTFNRIWRVXXQRPWM
      orf144a  130
      NTFNRIWRVNSQRPWMMQFLVYWALLTFGP LSLGVGISFXVGSVQDAALASGAPQWSGAL

60

```



130 140 150 160 170 180

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

```

      1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
      51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
5      101  CGGCGGCAAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCCTGCTG
      151  ACCGTGATGG TGGCGGTGCG TCGATTTCG CCGGTGTTTC ACCGNTGGTC
      201  GGATTCTGTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
      251  ACATGGTNTT CGACTATATC AATGCGTTCC GCAGACAGGC GAACCGGCTG
      301  ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCNGA TGCTGATTCC
10     351  GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
      401  CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
      451  CTGTCTTTGG GCGTGGGCAT TTCTTTATN GTCGGCTCGG TACAGGATGC
      501  CGCGCTTGCC TCAGGTGCGC CGCAGTGGTG GGGCGCGTTG CGAACGGCGG
      551  CGACGCTGAN CTTTCATGACG CTTTGTGCTG GGGGGCTGTA CCGCTNCGTG
15     601  CCAAACCGCT TCGTTCCTGC GCGGCANGCG TTTGTGCGGG CTTTGGCAAC
      651  AGCGTTCTGT CTGGAACCGC CGCGTTCCTT CTTTACTTGG TATATGGGCA
      701  ATTCGACGG CTACCGCTCG ATTTACGGNG CGTTTGCCGC CGTGCCGTTT
      751  TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
20     801  GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGNCT
      851  TCGACTCGCG CGGACGGTTT GACGACGTGT TGAATAACCT GCTGCTTCTG
      901  GATGCGGCGC AAAAGAAGG CNAAGCCTTG CCTGTTCAGG AGTTCAGACG
      951  GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
25    1001  CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAACACG
      1051  GGGCGGGATT CGATTGAGTT GAACGAATC TTCAAGCTCT TCGTTTACCG
      1101  TCCGTTGCCT GTGGAAGAGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
      1151  TGATGCCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
      1201  CAGGCGAAAA AACAGCAGCA ATCTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 624>:

```

      1  MTFLLQRLQGL ADNKICAFAW FVRRFDEER VPQAAASMTF TLLALVPVL
30     51  TVMVAVASIF PVFDRWSDSF VSFVNQTIVP QGADMVFDYI NAFREQANRL
      101  TAIGSVMLV TSXMLIRTID NTFNRIWRVN SQRPMWMOFL VYWALLTFGP
      151  LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRXV
20     201  PNRFVPARXA FVGALATAFC LETARSLFTW YMGNFDGYRS IYGAFAAVPE
      251  FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRFDSRGRF DDVLKILLLL
35     301  DAAQKEGXAL PVQEFRRHIN MGYDELGELL EKLARHGYIY SGRQGWVLKT
      351  GADSIENEL FKL FVYRPLP VERDHVNQAV DAVMMPCLOT LNMTLAEFDA
      401  QAKKQQS*

```

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

```

40     orf144a.pep  MTFLLQRLQGLADNKICAFAWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVAVASIF
      orf144-1     MTFLLQRLQGLADNKICAFAWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVAVASIF

      orf144a.pep  PVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSTXMLIRTID
45     orf144-1     PVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSTXMLIRTID

      orf144a.pep  NTFNRIWRVNSQRPMWMOFLVYWALLTFGP LSLGVGISFXVGSVQDAALASGAPQWSGAL
50     orf144-1     NTFNRIWRVNSQRPMWMOFLVYWALLTFGP LSLGVGISFVGSVQDAALASGAPQWSGAL

      orf144a.pep  RTAATLXFMTLLLWGLYRXVFNRFVPARXAFVGALATAFCLETARSLFTWYMGNFDGYRS
      orf144-1     RTAATLTFMTLLLWGLYRFVFNRFVPARQAFVGALATAFCLETARSLFTWYMGNFDGYRS

55     orf144a.pep  IYGAFAAVPPFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRXFSRGRFDDVLKILLLL
      orf144-1     IYGAFAAVPPFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL

      orf144a.pep  DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIENEL
60     orf144-1     DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIENEL

      orf144a.pep  FKL FVYRPLPVERDHVNQAVDAVMMPCLOT LNMTLAEFDAQAKKQQS 408
65     orf144-1     FKL FVYRPLPVERDHVNQAVDAVMTPCLOT LNMTLAEFDAQAKKRQ 406

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

5	orf144.pep	MTFLQRLQGLADNKICAFVFWRRFDEERVQPXAAASMTFTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFVFWRRFSEERVQAAASMTFTLLALVPVLTVMVAVASIF	60
10	orf144.pep	PVFDRWSDSEFVSFVNQTI VXPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144ng	PVFDRWSDSEFVSFVNQTI VXPQADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144.pep	NTFNRIWRVXXQRPWM	136
15	orf144ng	NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSGGAQQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

	1	MTFLQCWQGS	ADNKICAFV	FVIRRFSEER	VPQAAASMTF	TLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTI V	QADMVFDYI	DAFRDQANRL
20	101	TAIGSVMLV	TSLMLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLLWGLYRFV
	201	PNRFVPARQA	FVGALITAF	LETARFLFTW	YMGNF DGYS	IYGAFVAVPF
	251	FLWLNLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT
25	351	GADSIELSEL	FKLFVYRPLP	VERDHVNQAV	DAVMT PCLQT	LNMTLAEFDA
	401	QAKKQQS*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

	1	ATGACCTTTT	TACAACGTTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
	51	ATTTGCATGG	TTCGTCATCC	GCCGTTTCAG	TGAAGAGCGC	GTACCGCAGG
30	101	CAGCGGCGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTA CTG
	151	ACCGTAATGG	TCGCGGTCGC	TTCGATTTTC	CCCGTGTTTC	ACCGCTGGTC
	201	GGATTCGTTT	GTCTCCTTCG	TCAACCAAAC	CATTGTGCCG	CAGGGCGCGG
	251	ATATGGTGTT	CGACTATATC	GACGCATTCC	GCGATCAGGC	AAACCGGCTG
	301	ACCGCCATCG	GCAGCGTGAT	GCTGGTCGTA	ACCTCGCTGA	TGCTGATTCTG
35	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	GCGGGTTAAC	ACGCAACGCC
	401	CTTGATGAT	GCAGTTCTCT	GTTTATTGGG	CGTTGCTGAC	TTTCGGGCCT
	451	TTGCTTTTGG	GTGTGGGCAT	TTCCTTTATG	GTGCGGTCGG	TTCAAGACTC
	501	CGTACTCTCC	TCCGGAGCGC	AACAATGGGC	GCACGCGTTG	AAGACGGCGG
	551	CAAGGCTGGC	TTTCATGACG	CTTTTGCTGT	GGGGGCTGTA	CCGCTTCGTG
40	601	CCCAACCGCT	TCGTGCCCGC	CCGGCAGGCG	TTTGTCCGAG	CTTTGATTAC
	651	GGCATTTCTG	CTGGAGACGG	CACGTTTCCT	GTTCACTTGG	TATATGGGCA
	701	ATTTGACGCG	CTACCGCTCG	ATTACGGCG	CATTGCGCG	CGTGCCGTTT
	751	TTCTGCTGT	GGTTAAACCT	GCTGTGGACG	CTGGTCTTGG	GCGGGGCGGT
	801	GCTGACTTCG	TCGCTGTCTT	ATTGGCAGGG	CGAGGCCTTC	CGCAGGGGAT
45	851	TCGACTCGCG	CGGACGGTTT	GACGACGTGT	TGAAAATCCT	GCTGCTTCTG
	901	GATGCGGCGC	AAAAAGAAGG	CCGAACCTTG	TCCGTTTCAAG	AGTTCAGACG
	951	GCATATCAAT	ATGGGTACG	ATGAATTGGG	CGAGCTTTTG	GAAAAGCTGG
	1001	CGCGGTACGG	CTATATCTAT	TCCGGCAGAC	AGGGCTGGGT	TTTGAAAACG
	1051	GGGGCGGATT	CGATTGAGTT	GAGCGAACTC	TTCAAGCTCT	TCGTGTACCG
50	1101	CCCGTTGCct	gtggaAAGGG	ATCATGTGAA	CCAAGCTGtc	gaTGCGGTAA
	1151	TGAcgccgtG	TTTGACGACT	TTGAACATGA	CGCTGGCGGA	GTTTGACGCT
	1201	CAGgcgAAAA	AACAGCAGCA	GTCTTGA		

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-1>:

	1	MTFLQWQGL	ADNKICAFV	FVIRRFSEER	VPQAAASMTF	TLLALVPVL
55	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTI V	QADMVFDYI	DAFRDQANRL
	101	TAIGSVMLV	TSLMLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLLWGLYRFV
	201	PNRFVPARQA	FVGALITAF	LETARFLFTW	YMGNF DGYS	IYGAFVAVPF
	251	FLWLNLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
60	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT

351 GADSIELSEL FKLEFYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA  
401 QAKKQQQS\*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

```

5      orf144ng-1.pep MTFLLRWQGLADNKICAFWVIRRFSEERVPOAAAASMTFTLLALVPVLTVMVAVASIF
      orf144-1       MTFLLRWQGLADNKICAFWVIRRFSEERVPOAAAASMTFTLLALVPVLTVMVAVASIF

      orf144ng-1.pep PVFDRWSDSFVSFVNQTIVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
10     orf144-1       PVFDRWSDSFVSFVNQTIVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID

      orf144ng-1.pep NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
      orf144-1       NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL
15     orf144ng-1.pep KTAARLAFMTLLLLWGLYRFVFNRFVPAQAFVQALITAFCLLETARFLFTWYMGNFDDGYRS
      orf144-1       RTAATLTFMTLLLLWGLYRFVFNRFVPAQAFVQALITAFCLLETARSLFTWYMGNFDDGYRS

      orf144ng-1.pep IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGREFDDVLKILLLL
20     orf144-1       IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGREFDDVLKILLLL

      orf144ng-1.pep DAAQKEGRTLSVQEFRRHINMGYDELGELLEKLARYGYIYSGRQGWVLKTGADSIELSEL
25     orf144-1       DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYYIYSGRQGWVLKTGADSIELNEL

      orf144ng-1.pep FKLEFYRPLPVERDHVNQAVDAVMTPLCLQTLNMTLAEFDAQAKKQQQS
30     orf144-1       FKLEFYRPLPVERDHVNQAVDAVMTPLCLQTLNMTLAEFDAQAKKRQ

```

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 35 Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

```

1      ..AGACACGCC GCGCATCCG CATCGACACC GCCATCAACC CCGAAGTGA
51     AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGCTTC CTCTGGCTCA
101    GCACCGATAT CCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC
40     151     ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TCGCCAAAG
      201     CCTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

1      ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
51     TRRKWLDAHE RQHLRQSLLE TREHG*

```

45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

1      ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
51     CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGCTCCG CTCTGGCGGGG
101    CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
50     151     GAGTGGATAG GGATGACCGT CTTCGTCGTC CTCGGCATGC TCCAGTTTCA
      201     AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
      251     GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC
      301     GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
      351     CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGCTGA
401    CGATGTGTAT GTCATCGGC GACAACGGCA GCGAATGGCT CGACAGCGGA
55     451    CTCATGCGCG CCATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC

```

-352-

501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG  
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC  
 601 AGGCGCATGA CCCGCGAACG CCTCGAGGAG AACATGGCGA AAATGCGCCA  
 651 AATCAACGCA CGCATGGTCA AAAGCCGCGC CCATCTCGCC GCCACATCGG  
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCCCAC  
 751 CGTAAAATCG TCAACACCAAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT  
 801 GCAATCTCCC AAACCTCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC  
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA  
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG  
 1101 CCTGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

15 1 MNTSQNRNLV SRWLSNYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG  
 51 EWIGMTVFV LGMLOFQGA IYSKAVERMLG TVIGLGAGLG VLWLNQHYFH  
 101 GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWLD SG  
 151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG  
 201 RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH  
 251 RKIVNTTELL LTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING  
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR  
 351 TRRKWLDAHE RQHLRQSLLE TREHG\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N. meningitidis*:

					10	20	30
	orf146.pep				RHARRIRIDTAINPELEALAEHLHYQWQGF		
30	orf146a	KLNGSEIRLLDRHFTLLQTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWQGF					
		280	290	300	310	320	330
		40	50	60	70		
35	orf146.pep	LWLSTDMRQEISALVILLQRTRRKWLDAHERQHLRQSLLETREHGX					
	orf146a	LWLSTNMRQEISALVILLQRTRRKWLDAHERQHLRQSLLETREHSX					
		340	350	360	370		

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

40 1 ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA  
 51 CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCGGG CTCGGCGGGG  
 101 CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC  
 151 GAGTGGATAG GGATGACCGT CTTCGTCGTC CTCGGCATGC TCCAGTTTCA  
 201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTTCATCG  
 251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACAGCA TTATTTCAC  
 45 301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG  
 351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCGGGGCTGA  
 401 CGATGTGCAT GCTCATCGGC GACAACGGCA GCGAATGGTT CGACAGCGGC  
 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCCG  
 50 501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG  
 551 CCGACAACCT GACCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC  
 601 AGGCGCATGA CCCGCGAACG CCTCGAAGAG AACATGGCGA AAATGCGCCA  
 651 AATCAACGCA CGCATGGTCA AAAGCCGCGC CCACCTCGCC GCCACATCGG  
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCCCAC  
 751 CGTAAAATTG TCAACACCAAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT  
 801 GCAATCTCCC AAACCTCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC  
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA  
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 1051 ACCGCGCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG  
 1101 CCTGCTTGAA ACACGGGAAC ACAGTTGA

This encodes a protein having amino acid sequence <SEQ ID 634>:

```

      1  MNTSQNRNLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
    51  EWIGMTVFVV LGMLQFQGAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
   101  GNLLFYLTVG TASALAGWAA VGKNGYVPML AGLTMCMLIG DNGSEWFDSD
    5  151  LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLTDC SKMIAEISNG
   201  RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
   251  RKIVNTTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING
   301  RHARRIRIDT AINPELEALA EHLHYWQGF LWLSTNMRQE ISALVILLQR
   351  TRRKWLDAHE RQHLRQSLLE TREHS*

```

10 ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

```

      orf146a.pep  MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
      orf146-1     MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
   15  orf146a.pep  LGMLQFQGAISKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA
      orf146-1     LGMLQFQGAISKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA
   20  orf146a.pep  VGKNGYVPMLAGLTMCLIGDNGSEWFDSDGLMRAMNVLIGAAIAIAAAKLPLKSTLMWR
      orf146-1     VGKNGYVPMLAGLTMCLIGDNGSEWFDSDGLMRAMNVLIGAAIAIAAAKLPLKSTLMWR
   25  orf146a.pep  FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
      orf146-1     FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
   30  orf146a.pep  AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
      orf146-1     AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
   35  orf146a.pep  RHARRIRIDTAINPELEALAEHLHYWQGF LWLSTNMRQEISALVILLQTRRKWLDAHE
      orf146-1     RHARRIRIDTAINPELEALAEHLHYWQGF LWLSTNMRQEISALVILLQTRRKWLDAHE
   35  orf146a.pep  RQHLRQSLLETREHSX
      orf146-1     RQHLRQSLLETREHG

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from *N.gonorrhoeae*:

```

      orf146.pep  RHARRIRIDTAINPELEALAEHLHYWQGF 30
      orf146ng   KLNSEIRLLDRHFTLLQTDLQQTAAINGRHARRIRIDTAINPELEALAEHLHYWQGF 364
   45  orf146.pep  LWLSTDMRQEISALVILLQTRRKWLDAHERQHLRQSLLETREHG 75
      orf146ng   LWLSTNMRQEISALVIPLQTRRKWLDAHERQHLRQSLLETREHG 409

```

50 An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

```

      1  MSGVRFPSA PIPSTDPPSG SLCFFTFPLQ TASDMNSSQR KRLSGRWLNS
      51  YERYRHRRLI HAVRLGGTVL FATALARLLH LQHGEWIGMT VFVVLGMLQF
   101  QGAIYSNAVE RMLGTVIGLG AGLGVLWLNQ HYFHGNLLFY LTIGTASALA
   55  151  GWAAVGKNGY VPMLAGLTMC MLIGDNGSEW LDSGLMRAMN VLIIGAAIAIA
   201  AAKLLPLKST LMWRFMLADN LADCSKMIAE ISNGRRMTRE RLEQNVMKMR
   251  QINARMVKSR SHLAATSGES RISPSMMEAM QHAHRKIVNT TELLTTAAK
   301  LQSPKLNSE IRLDRHFTL LQTDLQQTAA LINGRHARRI RIDTAINPEL
   351  EALAEHLHYQ WQGF LWLSTN MRQEISALVI PLQTRRKWL DAHERQHLRQ
   401  SLLETREHG*

```

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

-354-

```

      1 ATGAACTCCT CGCAACGCAA ACGCCTTTCC GgccGCTGGC TCAACTCCTA
    51 CGAACGCTac cGCCaccGCC GCCTCATACA TGCCGTGCGG CTCGGCggaa
   101 ccgtCCTGTT CGCCACCGCA CTCGCCCGgc tACTCCACCT CCAaacaggc
   151 gAATGGATAG GGAtgaCCGT CTTCTGTCGTC CTCGGCATGC TCCAGTTCCA
   201 AGGCgcgatt tActccaacg cgggtgGAacg taTGctcggg acggtcatcg
   251 ggctgGGCGC GGGTTTGGgc gTTTTATGGC TGAACCAGCA TTAttccac
   301 ggcaacCTcc tcttctacct gaccatcggc acggcaagcg cactggccgg
   351 ctGGGCGGCG GTCGGCAAAA acggctacgt ccctatgctg GCGGGGctgA
   401 CGATGTGCAT gctcatcggc gACAACGGCA GCGAATGGCT CGACAGCGGC
   451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCCGCCATCG CCATTGCCCG
   501 CGCAAACCTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
   551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
   601 AGGCGTATGA CGCGCGAAGC TTTGGAGCAG AATATGGTCA AAATGCGCCA
   651 AATCAACGCA CGCATGGTCA AAAGCCGCGC CCACCTCGCC GCCACATCGG
   701 CGGAAAGCCG CATCAGCCCC TCCATGATGG AAGCCATGCA GCACGCCAC
   751 CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
   801 GCAATCTCCC AAACCTCAACG GCAGCGAAAT CCGGCTGCTC GACCGCCACT
   851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGCCGCCCT CATCAACGGC
   901 AGACACGCCG GCCGATCCG CATCGACACC GCCATCAACC CCGAAGTGA
  1001 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
  1051 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCCTCATCCT GCTGCAACGC
  1101 ACCGCGCGCA AATGGCTGGA TGCCCACGAA CGCCAACACC TGCGCCAAAG
      CCTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

```

    1 MNSSQRKRLS GRWLSYERY RHRRLIHAVR LGGTVLFATA LARLLHLQHG
   51 EWIGMTVFVV LGMLQFQGA IYNAVERMLG TVIGLGAGLG VLWLNQHYFH
  101 GNLLFYLTIG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWLD SG
  151 LMRAMNVLIG AAIATAA AKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
  201 RRMTRERLEQ NMVKMRQINA RMVKSRS LLA ATSGESRISP SMMEAMQH AH
  251 RKIVNTTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTAA LING
  301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQ R
  351 TRRKWLDAHE RQHLRQSLLE TREHG*

```

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

```

 35 orf146-1.pep MNTSQRNRLVSRWLSYERYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
    orf146ng-1 MNSSQRKRLSGRWLSYERYRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVV

 40 orf146-1.pep LGMLQFQGA IYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
    orf146ng-1 LGMLQFQGA IYNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA

 45 orf146-1.pep VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIAIAA AKLLPLKSTLMWR
    orf146ng-1 VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIAIAA AKLLPLKSTLMWR

 50 orf146-1.pep FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRS LLAATSGESRISP
    orf146ng-1 FMLADNLADCSKMIAEISNGRRMTRERLEQNVMKMRQINARMVKSRS LLAATSGESRISP

 55 orf146-1.pep AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
    orf146ng-1 SMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTAA LING

 60 orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
    orf146ng-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE

 65 orf146-1.pep RQHLRQSLLE TREHG X
    orf146ng-1 RQHLRQSLLE TREHG X

```

Furthermore, ORF146ng-1 shows homology with a hypothetical *E.coli* protein:

```

 65 sp|P33011|YEEA_ECOLI_HYPOTHETICAL_40.0_KD_PROTEIN_IN_COBU-SBMC_INTERGENIC_REGION
    >gi|1736674|gnl|PID|d1016553 (D90838) ORF_ID:o348#20; similar to [SwissProt
    Accession Number P33011] [Escherichia coli] >gi|1736682|gnl|PID|d1016560 (D90839)
    ORF_ID:o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]

```

-355-

>gi|1788318 (AE000292) f352; 100% identical to fragment YEEA\_ECOLI SW: P33011 but has 203 additional C-terminal residues [Escherichia coli] Length = 352  
 Score = 109 bits (271), Expect = 2e-23  
 Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

5 Query: 20 YRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVVLGMLQFQGAIIYSNAVERML 79  
 YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+  
 Sbjct: 15 YRHYRIVHGTRVALAFLLTFLIIRLFTIPESTWPLVTMVVIMGPISFWGNVVPRAFERIG 74

10 Query: 80 GTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAAVGNKYVPMLAGLTMCMMLI 139  
 GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++  
 Sbjct: 75 GTVLGSILGLIALQLE---LISLPLMLVWCAAMFLCGWLALGKKPYQGLLIGVTLAIVV 131

15 Query: 140 GDNGSEWLDSGLMRAMNVLIGXXXXXXXXKLLPLKSTLMWRFMLADNLADCSKMIAEISN 199  
 G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +  
 Sbjct: 132 GSPTGE-IDTALWRSQDVLGSLLAMLFITGIWPQRAFIHWRIQLAKSLTEYNRVYQSASF 190

20 Query: 200 GRRMTRERLEQNVMVKMRQINARMVKSRSHLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259  
 + R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V  
 Sbjct: 191 PNLLERPRLESHLQKLL---TDAVKMRGLIAPASKETRIPKSIYEGIQTINRNVLVCMLEL 247

25 Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXAALINGRHARRIRIDTAINPEL 316  
 + LN ++R D AL G +N +  
 Sbjct: 248 QINAYWATRPSPHFVLLNAQKLR--DTQHM MQILLSLVHALYEGNPQPVFANTEKLNDV 305

Query: 317 EALAEHL--HYQWQ-----GFLWLSTNMQRQEISALVILLQRTTRK 354  
 E L + L H+ + G++WL+ ++ L L+ R RK  
 Sbjct: 306 EELRQLLNHHDLKVVTETPIYGYVLNMETAHQLELLSNLICRALRK 352

On the basis of this analysis, including the identification of several transmembrane domains in the  
 30 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

35 1 ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA  
 51 GGGCAAATC GTCAGTGTGC GCGAACACAA CGAACGGCAG ATGGCGGACA  
 101 AGATTGTCGG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTCCTCGAT  
 151 GCGGGTACGC CGGCCGTGTG CGACCCGGGC GCGAACTCG CCCGCCGCGT  
 201 GCGTGAGGCC GGGTTTAAAG TCGTTCCCGT CGTGGGCGCA AC.GCGGTGA  
 40 251 TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATTT TTATTTCAAC  
 301 GGTTTTGTAC CGCCGAAATC GGGAGAACGC AGGAAACTGT TTGCCAAATG  
 351 GGTGCGGGCG GCGTTTCCTA TCGTCATGTT TGAACGCCCG CACCGCATCG  
 401 GTGCAGCGCT TGCCGATATG GCGGAACGT TCCCCGAACG CCGATTAATG  
 451 CTGGCGCGCG AAATTACGAA AACGTTTGAA ACGTTCTTAA GCGGCACGGT  
 501 TGGGGAAATT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG  
 45 551 AGATGGTGTG GGTGCTTTAT CCGGCGCAGG ATGAAAAACA CGAAGGCTTG  
 601 TCCGAGTCCG CGCAAAACAT CATGAAAATC CTCACAGCCG AGCTGCCGAC  
 651 CAAACAGGCG GCGGAGCTTG CTGCCAAAT CACGGGCGAG GGAAAGAAAG  
 701 CTTTGTACGA T..

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

50 1 ..AEDTRVTAQL LSAYGIQGKL VSVREHNERQ MADKIVGYLS DGMVVAQVSD  
 51 AGTPAVCDPG AKLARRVREA GFKVVPVVG XAVMAALSV GVEGSDFYFN  
 101 GFVPPKSGER RKLFAKWRA AFPIVMFETP HRIGAAALDM AELFPERRLM  
 151 LAREITKTFE TFLSGTVGEI QTALSADGDQ SRGEMVLVLY PAQDEKHGEL  
 201 SESAQNIMKI LTAELPTKQA AELAAKITGE GKKALYD..

55 Further work revealed the complete nucleotide sequence <SEQ ID 641>:

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC  
 51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC  
 101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG  
 151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTGAGG GCAAACCTCGT

-356-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85

201 CAGTGTGCGC GAACACAACG AACGGCAGAT GCGGACAAG ATTGTGCGCT  
251 ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG  
301 GCGGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG  
351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA  
401 GCGTGGCCGG TGTGGAAGGA TCCGATTTTT ATTTCAACGG TTTGTACCG  
451 CCGAAATCGG GAGAACGCAG GAACTGTTT GCCAAATGGG TCGGGGCGGC  
501 GTTCTCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG  
551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GCGCGCGCAA  
601 ATTACGAAA CTTTGAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA  
651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTGTTG  
701 TGCTTTATCC GCGCGAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG  
751 CAAAACATCA TGAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC  
801 GGAGCTTGCT GCCAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC  
851 TGGCTCTGTC TTGAAAAAC AAATAG

15 This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT  
51 RVTAQLLSAY GIQGLVSVR EHNERQMA DK IVGYLSDGMV VAQVSDAGTP  
101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP  
151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRMLLARE  
201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA  
251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWN K\*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E. coli* (accession number U18997)

ORF147 and *E. coli* ORF286 protein show 36% aa identity in 237aa overlap:

25 Orf147: 1 AEDTRVTAQLLSAYGIQGLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPG 60  
AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG  
Orf286: 43 AEDTRHTGLLQHFGINARLFLALDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPG 102  
30 Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLFKAVVRA 120  
L R RE F + GF+P KS RR  
Orf286: 103 YHLVRTCREAGIRVVPPLPGPCAITALSAAGLP SDRFCYEGFLPAKSKGRRDALKAEAE 162  
35 Orf147: 121 AFPIVMFETPHRIGALADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALSADGD 179  
++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +  
Orf286: 163 PRTLIFYESTHRLDLSLEDIVAVLGESRYVVLARELTKTWETIHGAPVGE LLAWVKEDEN 222  
40 Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELAAKITGEGKKALY 236  
+ +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY  
Orf286: 223 RRGEMVLIV-EGHKAQEEDLPADALRTLALQLAELPLKKAALAAEIHGVKKNALY 278

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

45 orf147.pep 10 20 30  
AEDTRVTAQLLSAYGIQGLVSVREHNERQ  
orf75a TLYVVATPIGNLADITLRLAVLQKADIICAEDTRVTAQLLSAYGIQGLVSVREHNERQ  
20 30 40 50 60 70  
50 orf147.pep 40 50 60 70 80 90  
MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSLVA  
orf75a MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSLVA  
80 90 100 110 120 130  
55 orf147.pep 100 110 120 130 140 150  
GVEGSDFYFNGFVPPKSGERRKLFKAVVRAAFPIVMFETPHRIGALADMAELFPERRLM  
orf75a GVAGSDFYFNGFVPPKSGERRKLFKAVVRAAFPIVMFETPHRIGATLADMAELFPERRLM  
140 150 160 170 180 190  
60 orf147.pep 160 170 180 190 200 210  
LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI



10 ORF147a is identical to ORF75a, which includes aa 56-292 of ORF75.

ORF147 shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) from *N. gonorrhoeae*:

An ORF147ng nucleotide sequence <SEQ ID 643> was predicted to encode a protein having amino acid sequence <SEQ ID 644>:

Further work revealed the following gonococcal DNA sequence <SEQ ID 645>:

45	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCTGCG
	101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATTTGTGC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTCAGG	GCAGGTTGGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGCAAG	GTAATTGCGTT
50	251	TCCTTTCAGA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
	301	GCCGTGTGCG	ACCCGGGGCG	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAAG	CGCGGTAATG	CGCGCGTTGA
	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG
	451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGCGGC
55	501	ATTTCTGTGC	GTCATGTTTG	AAACGCCGCA	CCGAATCGGG	GCAACGCTTG
	551	CCGATATGGC	GGAAATGTTC	CCGAAACGCC	GTCTGATGCT	GGCGCGCGAA
	601	ATCAGAAAA	CGTTTGAAC	GTTCTTAAGC	GGCACGGTTG	GGGAAATTCA
	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTTCG
	701	TGCTTTATCC	GGCGCAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCTGCG
60	751	CAAAATGCGA	TGAAAATCCT	TGCGGCCGAG	CTGCCGACCA	AGCAGGCGGC
	801	GGAGCTTGCC	GCCAAGATTA	CAGGTGAGGG	CAAAAAGGCT	TTGTACGATT
	851	TGGCACTGTC	GTTGGAAAAAC	AAATGA		

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

      1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
    51 RVTAQLLSAY GIQGRIVSVR EHNERQMADK VIGFLSDGLV VAQVSDAGTP
   101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
    5 151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLADMAELF FERRMLLARE
   201 ITKTFFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
   251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

```

ORF147ng shows homology to a hypothetical *E.coli* protein:

```

10 sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
   (F286)
   >gi|606086 (U18997) ORF_f286 [Escherichia coli]
   >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
   [Escherichia coli] Length = 286
   Score = 218 bits (550), Expect = 3e-56
15 Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

Query: 4 KHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
      K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
Sbjct: 2 KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAAEDTRHTGLLQLHFGIN 59

20 Query: 64 GRLVSVREHNERQMADKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
      RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
Sbjct: 60 ARLFALHDHNEQKKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVVRTCREAGIRVVPL 119

25 Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFKAWVRAAFPVVMFETPHRIGATL 183
      G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
Sbjct: 120 PGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAI EAEPRTLIFYESTHRLDLSL 179

30 Query: 184 ADMAELFPERR-LMLAREITKTFFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
      D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
Sbjct: 180 EDIVAVLGESRYVVLARELTKTWTIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238

Query: 243 HEGLSESAQNAMKILAAELPTKQAAELA AKITGEGKKALYDLAL 286
      E L A + +L AELP K+AA LAA+I G K ALY AL
35 Sbjct: 239 EEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALYKYAL 282

```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 77

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 647>

```

      1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
      51 AACCGGTCGC ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTGCT
   101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC AACTTATTT CGGCATCAAC
   45 151 TACCAATACT ATCGCGACTT TGCCGAAAT AAAGGCAAGT TTGCAGTCGG
      201 GGCGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
      251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
      301 GTGGCGGCAT TGGTGGGCGT ATCAATATAT TGTGAGCGTG GCACATAACG
      351 GCGGCTATAA CAACGTTGAT TTTGGTGCGG AAGGAak.AA tATCCC.GAT
      401 CAACAACGww TTACTTATAA AATTGTGAAA CGGAATAATT ATAAAGCAGG
   50 451 GACTAAAGGC CATCCTTATG GCGGCGATTA TCATATGCCG CGTTTGCCATA
      501 AATwTGTCAC AGATGCAGAA CCTGTTGAAA TGACCACTTA TATGGATGGG
      551 CGGAAATATA TCGATCAAAA TAATTACCCT GACCGTGTTC GTATTGGGGC
      601 AGGCAGGCAA TATTGGCGAT CTGATGAAGA TGAGCCCAAT AACCGCGAAA
      651 GTTCATATCA TATTGCAAGT .....
   55 701 ..... GGCTC ACCAATGTTT ATCTATGATG CCCAAAAGCA
      751 AAAGTGGTTA ATTAATGGGG TATTGCAAAC GGGCAACCCC TATATAGGAA
      801 AAAGCAATGG CTTCCAGCTG GTTCGTAAAG ATTGGTTCTA TGATGAAATC
      851 TTTGCTGGAG ATACCCATTC AGTATTCTAC GAACCACTC AAAATGGGAA
      901 ATACTCTTTT AACGACGATA ATAATGGCAC AGGAAAAATC AATGCCAAAC

```

-359-

5 951 ATGAACACAA TTCTCTGCCT AATAGATTAA AAACACGAAC CGTTCAATTG  
 1001 TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC  
 1051 AGGTGGTGTC AACAGTTATC GACCCAGACT GAATAATGGA GAAAATATTT  
 1101 CCTTTATTGA CGAAGGAAAA GGCGAATTGA TACTTACCAG CAACATCAAT  
 1151 CAAGGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCCTGA  
 1201 AAATAACGAA ACTTGGCAAG GCGCGGGCGT TCAATACAGT GAAGACAGTA  
 1251 CCGTTACTTG GAAAGTAAAC GGCGTGCGAA ACGACCGCCT GTCCAAAATC  
 1301 GGCAAAGGCA CGCTG.....  
 //  
 10 2101 ..... GATAAAG  
 2151 TGACTGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC  
 2201 GATCAGGCTC ATTTAAATCT CACAGGGCTT GCCACACTCA ACGGCAATCT  
 2251 TAGTGCAAAT GGCGATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA  
 2301 ACGGCAACCK TAGCCTCGtG G.sAATGcCC AAGCAACATT TAATCAAGCC  
 15 2351 ACATTAACG GCAACACATC GGCTTCgGGC AATGCTTCAT TTAATCTAAG  
 2401 CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCGGC AACGCTAAGG  
 2451 CAAACGTAAG CCATTCCGCA CTCAACGGTA ATGTCTCCCT AGCCGATAAG  
 2501 GCAGTATTC ATTTTGAAAG CAGCCGCTT ACCGACAAA TCAGCGGCGG  
 2551 CAaGATACG GCATTACACT TAAAAGACAG CGAATGGACG CTGCGCTCag  
 20 2601 GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACaCTCAAT  
 2651 TCGCCTATC GCCACGATGC GGCAGGGGCG CAAACCGGCA GTGCGACAGA  
 2701 TCGCGCGCGC CGCCGTTGCG GCGGTTGCGC CCGTTCCTTA TTATmCGTTA  
 2751 CACCGCCAAC TTCGGTAGAA TCCCGTTCA ACACGCTGAC GGTAACCGTC  
 2801 AAATTGAACG GTCAGGGAAC ATTCCGCTTT ATGTGCGAAC TCTTCGGCTA  
 25 2851 CCGCAGCGAC AAATTGAAGC TGGCGGAAAG TTCCGAAGGC ACTTACACCT  
 2901 TGGCGGTCAA CAATACCGGC AACGAACCTG CAAGCCTCGA ACAATTGACG  
 2951 GTAGTGGAAG GAAAAGACAA CAAACCGCTG TCCGAAAACC TTAATTTAC  
 3001 CCTGCAAAAC GAACACGTCG ATGCAGGCGC GTGG.....  
 //  
 30 3551 ..... TTAGAC CGCGTATTG CCGAAGACCG  
 3601 CCGCAACGCC GTTTGGACAA GCGGCATCCG GGACACCAA CACTACCGTT  
 3651 CGCAAGATT CCGCGCCTAC CGCAACAAA CCGACCTGCG CCAAATCGGT  
 3701 ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC GGCATCCTGT TTTGCGACAA  
 35 3751 CCGGACCGAA AACACCTTCG ACGACGGCAT CGGCAACTCG GCACGGCTTG  
 3801 CCCACGGCGC CGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC  
 3851 ATCAGCGCG GCGCGGGTT TTAGCAGCG CAGCCTTTcA GACGGCATCG  
 3901 GAGsmAAAwT CCGCGCGCG GTGctGCATT ACGGCATTCA GGCACGAtAC  
 3951 CGCGCGGgtt tCgGCGgAtt CGGCATCGAA CCGCACATCG GCGCAACGCG  
 40 4001 ctATTTCTGTC CAAAAAGCG ATTACCGCTA CGAAAACGTC AATATCGCCA  
 4051 CCCCCGGCCT TGCATTCAAC CGcTACCGCG CCGGCATTaA GGCAGATTAT  
 4101 TCATTCAAAC CGGCGCAACA CATTTCCATC ACGCCTTATT TGAGCCTGTC  
 4151 CTATACCGAT GCCCTTCGG GCAAAGTCCG AACACGCGTC AATACCGCCG  
 4201 TATTGGCTCA GGATTTCGGC AAAACCCGCA GTGCGGAATG GGgCGTAAAC  
 45 4251 GCCGAATCA AAGTTTTCAC GCTGTCCCTC CACGCTGCCG CCGCAAAGG  
 4301 CCCGCAACTG GAAGCGCAAC ACAGCGCGG CATCAAATTA GGCTACCGCT  
 4351 GTTAA...

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

50 1 MKTTDKRTE THRKAPKTGR IRFXAAYLAI CLSFGILPQA WAGHTYFGIN  
 51 YQYYRDFAE NKGKFAVGA KD IEVYNKKGEL VGKSMTRKAPM IDFSVVS RNG  
 101 VAALVG VQYI VSVAHNNGYN NVDFGAEGXN IXDQXRXYK IVKRNNYKAG  
 151 TKGHPYGGDY HMPRLHKXVT DAEPVEMTSY MDGRKYIDQN NYPDRVRIGA  
 201 GRQYWRSD EEPNNRESSYH IAS.....GS PMFIYDAQKQ  
 251 KWLINGVLQT GNPYIGKSNG FQLVRKDWFY DEIFAGDTHS VFYEPRQNGK  
 301 YSFNDNNGT GKINAKHEHN SLPNRLKTRT VQLFNVSLS E TAREPVYHAA  
 55 351 GGVNSYRPR L NNGENISFID EGKGELILTS NINQGAGGLY FQGDFTVSPE  
 401 NNETWQAGV HISEDSTVTW KVN GVANDRL SKIGKGL..  
 //  
 701 .....DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL  
 751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS  
 801 DHAVQNGSLT LSGNAKANVS HSA LNGNVSL ADKAVFHFES SRFTGQISGG  
 851 KDTALHLKDS EWTLP SGXEL GNLNL DNATI TLNSAYRHDA AGAQTGSATD  
 901 APRRRSRRSR RSLXVTPPT SVESRFNTLT VNGKLNGQGT FRFMSELFY  
 951 RSDKLKLAES SEGTYTLAVN NTGNEPASLE QLTVVEGKDN KPLSENLNFT  
 65 1001 LQNEHVDAGA W.....  
 //  
 1151 ..... LDRVFAEDR  
 1201 RNAVWTS GIR DTKHYRSQDF RAYRQOTDLR QIGMQKNLGS GRVGILFSHN  
 1251 RTENTFDDGI GNSARLAHGA VFGQY GIDRF YIGISAGAGF SSGSLSDGIG  
 1301 XKRRRRVLHY GIQARYRAGF GFGIEPHIG ATRYFVQKAD YRYENVNIAT  
 70 1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDAAAG KVRTRVN TAV

1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW  
 1451 \*

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

```

5      1  ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
      51  AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCCTGTCT
101    101  TCGGCATTCT TCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
      151  TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
      201  GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
      251  CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
10      301  TGGCGGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
      351  CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGAAAT CCCGATCAAC
      401  ATCGTTTTAC TTATAAAATT GTGAAACGGA ATAATTATAA AGCAGGGACT
      451  AAAGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCATAAATT
15      501  TGTACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA
      551  AATATATCGA TCAAATAAAT TACCCTGACC GTGTTTCGTAT TGGGGCAGGC
      601  AGGCAATATT GGCATCTGA TGAAGATGAG CCCAATAACC GCCAAAGTTC
      651  ATATCATATT GCAAGTGCGT ATTCTTGGCT CGTTGGTGGC AATACCTTTG
      701  CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGTGTAG TGAAAAAATT
      751  AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
20      801  TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
      851  ATGGGGTATT GCAAACGGGC AACCCTATA TAGGAAAAAG CAATGGCTTC
      901  CAGCTGGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
      951  CCATTTCAGTA TTCTACGAAC CACGTCAAAA TGGGAAATAC TCTTTTAACG
25     1001  ACGATAATAA TGGCACAGGA AAAATCAATG CCAAACATGA ACACAATTCT
      1051  CTGCCTAATA GATTA AAAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT
      1101  ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGTGTCAACA
      1151  GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTCT TATTGACGAA
      1201  GGAAAAGGCG AATTGATACT TACCAGCAAC ATCAATCAAG GTGCTGGAGG
      1251  ATTATATTTT CAAGGAGATT TTACGGTCTC GCCTGAAAAA AACGAAACTT
30     1301  GGCAAGGCGC GGGCGTTTCA ATCAGTGAAG ACAGTACCGT TACTTGAAA
      1351  GTAAACGGCG TGGCAAACGA CCGCCTGTCC AAAATCGGCA AAGGCACGCT
      1401  GCACGTTCAA GCCAAAGGGG AAAACCAAGG CTCGATCAGC GTGGGCGACG
      1451  GTACAGTCAT TTTGGATCAG CAGGCGACG ATAAAGGCAA AAAACAAGCC
      1501  TTTAGTAAA TCGCTTGGT CAGCGCGAGG GGTACGGTGC AACTGAATGC
35     1551  CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCGGCTTT CGCGGCGGAC
      1601  GTTTGGATTT AAACGGGCAT TCGCTTTCGT TCCACCGTAT TCAAAATACC
      1651  GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
      1701  TACCATTACA GGCAATAAAG ATATTGCTAC AACCGGCAAT AACAACAGCT
      1751  TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
40     1801  ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTTACCAGC CCGCCGACAG
      1851  AGACCGCACCC CTGCTGCTTT CCGGCGGAAC AAATTTAAAC GGCAACATCA
      1901  CGCAAACAAA CGGCAAACTG TTTTTCAGCG GCAGACCAAC ACCGCACGCC
      1951  TACAATCATT TAAACGACCA TTGGTTCGCA AAAGAGGGCA TTCTCGCGG
45     2001  GGAAATCGTG TGGGACAACG ACTGGATCAA CCGCACATTT AAAGCGGAAA
      2051  ACTTCCAAAT TAAAGGCGGA CAGGCGGTGG TTTCCCGCAA TGTGTCGCAA
      2101  GTGAAAGGCG ATTGGCATTT GAGCAATCAC GCCCAAGCAG TTTTGGTGT
      2151  CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTTCGAC TGGACGGGTC
      2201  TGACAAATTG TGTCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
50     2251  TTGACTAAGA CCGACATCAG CGGCAATGTC GATCTTGCCG ATCAGCTCA
      2301  TTTAAATCTC ACAGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAAATG
      2351  GCGATACACG TTATACAGTC AGCCACAACG CCACCCAAAA CGGCAACCTT
      2401  AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
      2451  CAACACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGCCG
55     2501  TACAAAACGG CAGTCTGACG CTTTCCGGCA ACGCTAAGGC AAACGTAAGC
      2551  CATTCGCGAC TCAACGGTAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
      2601  TTTTGAAAGC AGCCGCTTTA CCGGACAAAT CAGCGGCGGC AAGGATACGG
      2651  CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTGAGG CACGGAATTA
      2701  GGCAATTTAA ACCTTGACAA CGCCACCATT ACACTCAATT CCGCCTATCG
      2751  CCACGATGCG GCAGGGGCGC AAACCGGCAG TGCGACAGAT GCGCGCGGCC
60     2801  GCCGTTGCGC CCGTTCGCGC CGTTCCTTAT TATCCGTTAC ACCGCCAACT
      2851  TCGGTAGAAAT CCCGTTTCAA CACGCTGACG GTAAACGGCA AATTGAACGG
      2901  TCAGGGAACA TTCCGCTTTA TGTCGGAACCT CTTGCGCTAC CGCAGCGACA
      2951  AATTGAAGCT GCGGGAAGT TCCGAAGGCA CTTACACCTT GGCGGTCAAC
      3001  AATACCGGCA ACGAACCTGC AAGCCTCGAA CAATTGACGG TAGTGGAAGG
65     3051  AAAAGACAAC AAACCGCTGT CCGAAAACCT TAATTTCAAC CTGCAAAACG
      3101  AACACGTCGA TGCCGGCGCG TGCGGTTACC AACTCATCCG CAAAGACGGC
      3151  GAGTTCCGCC TGCATAATCC GGTCAAAGAA CAAGAGCTTT CCGACAACCT
      3201  CGGCAAGGCA GAAGCCAAAA AACAGGCGGA AAAAGACAAC GCGCAAAGCC
      3251  TTGACGCGCT GATTGCGGCC GGGCGGATG CCGTCGAAAA GACAGAAAGC
70     3301  GTTGCCGAAC CGGCCCGGCA GGCAGGCGGG GAAAATGTCG GCATTATGCA

```

-361-

5  
 10  
 15  
 20

```

3351 GCGGAGGAA GAGAAAAAAC GGGTGCAGGC GGATAAAGAC ACCGCCTTGG
3401 CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCACCGC CTTCCCCCGC
3451 GCCCGCCGCG CCCCGCGGGA TTTGCCGCAA CTGCAACCCC AACCGCAGCC
3501 CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAATAGC GGTTTGAGTG
3551 AATTTTCCGC CACGCTCAAC AGCGTTTTCG CCGTACAGGA CGAATTAGAC
3601 CGCGTATTTC CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG
3651 GGACACCAA CACTACCGTT CGCAAGATT CCGCGCCTAC CGCCAACAAA
3701 CCGACCTGCG CCAAATCGGT ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC
3751 GGCATCCTGT TTTTCGCACAA CCGGACCGAA AACACCTTCG ACGACGGCAT
3801 CGGCAACTCG GCACGGCTTG CCCACGGCGC CGTTTTCGGG CAATACGGCA
3851 TCGACAGGTT CTACATCGGC ATCAGCGCGG GCGCGGGTTT TAGCAGCGGC
3901 AGCCTTTCAG ACGGCATCGG AGGCAAAATC CGCCGCCGCG TGCTGCATTA
3951 CGGCATTTCAG GCACGATACC GCGCCGTTT CCGCGGATTC GGCATCGAAC
4001 CGCACATCGG CGCAACGCGC TATTTTCGTC AAAAAGCGGA TTACCGCTAC
4051 GGAACGTC AATATCGCAC CCCCGCCTT GCATTCAACC GCTACCGCGC
4101 GGGCATTAA GAGATTATT CATTCAAACC GCGCAACAC ATTTCCATCA
4151 CGCCTTATTT GAGCTGTGCC TATACCGATG CCGCTTCGGG CAAAGTCCGA
4201 ACACGCGTCA ATACCGCCGT ATTGGCTCAG GATTTCGGCA AAACCCGCAG
4251 TCGGGAATGG GGCGTAAACG CCGAAATCAA AGTTTTCACG CTGTCCCTCC
4301 ACGCTGCCGC CGCCAAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC
4351 ATCAAATTAG GCTACCGCTG GTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

25  
 30  
 35  
 40  
 45  
 50

```

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
51 YQYYRDFAE NKGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVS RNG
101 VAALVG DQYI VSAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
151 KGHYPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
201 RQYWRSEDE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI
251 KHSFYGF LPT GGSFGDSGSP MFIYDAQKQK WLVINGVLOTG NPYIGKSNGF
301 QLVKRDW FID EIFAGDTHSV FYEPRQNGKY SFNDDNNGTG KINAKHEHNS
351 LPNRLKTRTV QLFNVLSLET AREPVYHAAG GVNSYRPRLN NGENISFIDE
401 GKGEILITSN INQAGAGGLYF QGDFTVSPEN NETWQAGGVH ISEDSTVTWK
451 VNGVANDRLS KIGKGT LHVQ AKGENQGSIS VGDGTVILDQ QADDKGKKQA
501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGRDLNLNGH SLSFHRIQNT
551 DEGAMIVNHN QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNWGFGEKD
601 TTKTNGRLNL VYQPAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
651 YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK
701 VKGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
751 LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL
801 SLVGNAQATF NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
851 HSA LNGNVS LADKAVFHES SRFTGQISGG KDTALHLKDS EWTLPSTEL
901 GNLNL DNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT
951 SVESR FNTLT VNGKLNQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN
1001 NTGNEPASLE QLTVEGKDN KPLSENLNFT LQNEHVDAGA WRYQLIRKDG
1051 EFRLHNPVKE QELSDKLGA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
1101 VAEPARQAGG ENVGIMQAE EKKRQVADKD TALAKQREAE TRPATTAFFR
1151 ARRARRDLPO LQPQPQPQPQ RDLISRYANS GLSEFSATLN SVFAVQDELD
1201 RVFAEDRRNA VWTSGIRDTK HYRSQDFRAY RQQTDLRQIG MQKNLGSGRV
1251 GILFSHNRTE NTFDDGICNS ARLAHGAVFC QYGIDRFYIG ISAGAGFSSG
1301 SLSDGIGGKI RRRVLHYGIQ ARYRAGFGGF GIEPHIGATR YFVQKADYRY
1351 ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDAAAGKVR
1401 TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAGK PQLEAQHSAG
1451 IKLGYRW*
  
```

Computer analysis of these sequences gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

60

```

      10      20      30      40      50      60
orfl.pep MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAE
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orfla    MKTTDKRTTETHRKAPKTGRIRFSPAYLAI CLSFGILPQAWAGHTYFGINYQYYRDFAE
      10      20      30      40      50      60

      70      80      90     100     110     120
orfl.pep KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVGVDQYIVSAHNGGYN
  
```

orf1.pep

orf1a

10

15

orf1.pep

20

orf1.pep

25

30

orf1.pep

35

40

orf1.pep

45

orf1.pep

50

55

0111.ppt

60

orf1.pep

65

orfla

70

490 500 510 520 530 540  
GDTRYTVSHNATONGNXSLVXNAOATENOATLNGNTSASGNASFNLSDHAVONGSITLGG

-363-

```

|||||:|||||
orfla  GDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSGNASFNLSNNAQNGSLTSLSD
      780      790      800      810      820      830

5      550      560      570      580      590      600
orfl.ppep  NAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGNL
orfla      NAKANVSHSALNGNVSLADKAVFHFENSRTGQLSGSKXTALHLKDSEWTLPSGTELGNL
      840      850      860      870      880      890

10      610      620      630      640      650      660
orfl.ppep  NLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRRSLLXVTPPTSVESRFNTLTVNG
orfla      NLDNATITLNSAYRHDAAGAQTGXVSDTPRRRSRRS---LLSVTPPTSVESRFNTLTVNG
      900      910      920      930      940      950

15      670      680      690      700      710      720
orfl.ppep  KLNGQGTFRFMSELFGYRSDKLLAESSEGTYYTLAVNNTGNEPASLEQLTVVEGKDNKPL
orfla      KLNXXQGTFRFMSELFGYRSDKLLAESSEGTYYTLAVNNTGNEPVSLDQLTVVEGKDNKPL
      960      970      980      990      1000     1010

20      730      740      750
orfl.ppep  SENLNFTLQNEHVDAGAW-----
orfla      SENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAEAKKQAEKDNAQS
      1020     1030     1040     1050     1060     1070

25      1080     1090     1100     1110     1120     1130
orfl.ppep  -----
orfla      LDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEEKRVQADKDSALAKQREAE TRP
      1080     1090     1100     1110     1120     1130

30      760
orfl.ppep  -----LDR
orfla      XTTFAPPRARXARRDLPPQPQPQPQPQORDLXSRYANSGLSEFSATLNSVFVAVQDELDR
      1140     1150     1160     1170     1180     1190

35      770      780      790      800      810      820
orfl.ppep  VFAEDRRNAVWTSIGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVIGILFSHNR TEN
orfla      VFAEDRRNAVWTSIXIRXTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVIGILFSHNR TEN
      1200     1210     1220     1230     1240     1250

40      830      840      850      860      870      880
orfl.ppep  TFDDGIGNSARLAHGA VFGQYQYIDRFYIGISAGAGFSSGSLSDGIGXKXRRRVLHYGIQA
orfla      XFDDGIGNSARLAHGA VFGQYQYIGRFDIGISTGAGFSSGSLSDGIGKIRRRVLHYGIQA
      1260     1270     1280     1290     1300     1310

45      890      900      910      920      930      940
orfl.ppep  RYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI
orfla      RYRAGFGGFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHX
      1320     1330     1340     1350     1360     1370

50      950      960      970      980      990      1000
orfl.ppep  SITPYLSLSYTDAA SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAA AKGP
orfla      SITPYXLSYTDAA SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHAAA AKGP
      1380     1390     1400     1410     1420     1430

55      1010     1020
orfl.ppep  QLEAQHSAGIKLGYRWX
orfla      QLEAQHSAGIKLGYRWX
      1440     1450

```

70 The complete length ORF1a nucleotide sequence &lt;SEQ ID 651&gt; is:

	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCGAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCTGCTTA	CTTAGCCATA	TGCCTGTCTG
	101	TCGGCATTCT	TCCCAAGCT	TGGGCGGGAC	ACACTTATTT	CGGCATCAAC
5	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGGTNT	ACAACAAAAA	AGGGGAGTTG	GTCGGCAAAAT
	251	CAATGACAAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTGTC	GCGTAACGGC
	301	GTGGCGGCAT	TGGTGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
	351	CGGCTATAAC	AACGTTGATT	TTGGTGCGGA	AGGAAGNAAT	CCCGATCAGC
10	401	ACCGTTTTTC	TTACCAAATT	GTGAAAAGAA	ATAATTATAA	GCCTGACAAT
	451	TCACACCCTT	ACAACGGCGA	TTANCATATG	CCGCGTTTGC	ATAAATTTGT
	501	CACAGATGCA	GAACCTGTCT	AAATGACGAG	TGACATGAGG	GGGAATACCT
	551	ATTCCGATAA	AGAAAAATAT	CCCGAGCGTG	TCCGCATCGG	CTCAGGACAC
	601	CACATATTGG	GTTATGATGA	TGACAAACAC	GGCGATTTAT	CCTACTCCGG
	651	CGCATGGTTA	ATTGGCGGCA	ATACACATAT	GCAGGGTTGG	GGAAAAATGT
15	701	CGGTANTTAG	TTTGAGCGGC	GATGTGCGCC	ATGCCAACGA	CTATGGCCCT
	751	ATGCCGATTG	CAGGTGCGGC	AGGCGACAGC	GGTTCGCCAA	TGTTTATTTA
	801	TGACAAAACA	AACAATAAAT	GGCTGCTCAA	CGGAGTTTTA	CAAACCGGCT
	851	ACCTTTATTC	CGGCAGGGAA	AACGGTTTCC	AGCTGATACG	CAAAGATTGG
	901	TTCTACGATG	ACATTTACAG	AGGCGATACA	CATACCGTCT	NTTTTGAACC
20	951	GCGCAGTAAC	GGACATTTTT	CCTTTACATC	CAACAACAAC	GGTACGGGTA
	1001	CGGTAAACAGA	AACCAACGAA	AAGGTNTCCA	ATCCAAAGCT	TAAAGTACAG
	1051	ACAGTCCGAC	TGTTTGACGA	ATCTTTGAAT	GAAACTGATA	AAGAACCAGT
	1101	TTACGCGGCA	GGGGGTGTTA	ATCAGTACCG	TCCAAGGTTA	AACAACGGTG
	1151	AAAACCTTTC	TTTTATCGAT	TACGGCAACG	GCAAACTCAT	CTTATCAAAC
25	1201	AACATCAACC	AAGGCGCGGG	CGGTTTGTAT	TTTGAAGGTG	ATTTTACGGT
	1251	CTCGCCTGAA	AACAACGAAA	CGTGGCAAGG	CGCGGGCGTT	CATATCAGTG
	1301	AAGACAGTAC	CGTTACTTGG	AAAGTAAACG	GCGTGGCAAA	CGACCCGCTG
	1351	TCCAAAATCG	GCAAAGCGAC	GCTGCACGTT	CAAGCCAAAG	GGGAAAAACCA
30	1401	AGGCTCGATC	AGCGTGGGCG	ACGGTACAGT	CATTTTGGAT	CAGCAGGCAG
	1451	ACGATAAAGG	CAAAAAACAA	GCCTTTAGTG	AAATCGGCTT	GNTCAGCGGC
	1501	AGGGGTACGG	TGCAACTGAA	TGCCGATAAT	CAGTTCAACC	CCGACAACT
	1551	CTATTTTCGG	TTTCGCGGCG	GACGTTTGGA	TTTAAACGGG	CATTGCTTTT
	1601	CGTTCCACCG	TATTCAAAAT	ACCGATGAAG	GGGCGATGAT	TGNCNATCAT
35	1651	AATGCCACAA	CAACATCCAC	CGTTACCATT	ACAGGGAATG	AAAGTATTAC
	1701	ACAACCGAGT	GGTAAGAATA	TCAATAGACT	TAATTACAGC	AAAGAAATTG
	1751	CCTACAACGG	TTGGTTTGCG	GAGAAAGATA	CGACCAAAAC	GAACGGGCGG
	1801	CTCAACCTTG	TTTACCAGCC	CGCCGCAGAA	GACCGCACCC	NGCTGCTTTC
	1851	CGGCGGAACA	AATTTAAACG	GCAACATCAC	GCAAAACAAAC	GGCAAACTGT
40	1901	TTTTCAGCGG	CAGACCGACA	CCGCACGCCT	ACAATCATT	AGGAAGCGGG
	1951	TGGTCAAAAA	TGGAAGGTAT	CCCACAAGGA	GAAATCGTGT	GGGACAACGA
	2001	CTGGATCNAC	CGCACGTTTA	AAGCGGAAAA	TTTCCATATT	CAGGGCGGGC
	2051	AGGCGGTGAT	TTCCCGCAAT	GTTGCCAAAG	TGGAAGGCGA	TTGNCATTTG
	2101	AGCAATCACG	CCCAAGCAGT	TTTTGGTGTC	GCACCGCATC	AAAGCCATAC
45	2151	AATCTGTACA	CGTTCCGACT	GGACNGGTCT	GACAAATTGT	GTCGAANAAA
	2201	NCATTACCGA	CGATAAAGTG	ATTGCTTCAT	TGACTAAGAC	NGACNTNAGC
	2251	GGCANTGTNA	GNCTNNCCNA	TNACGNTNNT	TNAAANCTCN	CNGGGCNTGC
	2301	NNCACTNAAN	GGCAATCTTA	GTGCAATGGA	CGATACACGT	TATACAGTCA
	2351	GCCACAACGC	CACCCAAAAC	GGCAACCTTA	GCCTCGTGGG	CAATGCCCAA
50	2401	GCAACATTTA	ATCAAGCCAC	ATTAAACGGC	AACNCATCGG	NTTCGGGCAA
	2451	TGCTTCATTT	AATCTAAGCA	ACAACGCCGC	ACAAAACGGC	AGTCTGACGC
	2501	TTTCCGACAA	CGCTAAGGCA	AACGTAAGCC	ATTCCGCACT	CAACGGCAAT
	2551	GTTCCCTTAG	CCGATAAGGC	AGTATTCCAT	TTTGAAAAAC	GCCGCTTTAC
	2601	CGGACAACTC	AGCGGCAGCA	AGGANACAGC	ATTACACTTA	AAAGACAGCG
55	2651	AATGGACGCT	GCCGTGAGGC	ACGGAATTAG	GCAATTTAAA	CCTTGACAAC
	2701	GCCACCATTA	CACTCAATTC	CGCCTATCGC	CACGATGCTG	CAGGCGCGCA
	2751	AACCGGCAGN	GTGTGAGACA	CGCCGCGCCG	CCGTTGCGCG	CGTTCCCTAT
	2801	TATCCGTTAC	ACCGCCAAC	TCGGTAGAAT	CCCGTTTCAA	CACGCTGACG
	2851	GTAACCGGCA	AATTGAACNG	TCAAGGAACA	TTCCGCTTTA	TGTCGGAAC
60	2901	CTTCGGCTAC	CGAAGCGACA	AATTGAAGCT	GGCGGAAAGT	TCCGAAGGNA
	2951	CTTACACCTT	GGCGGTCAAC	AATACCGGCA	ACGAACCCGT	AAGCCTCGAT
	3001	CAATTGACGG	TAGTGAAGG	GAAAGACAAC	AAACCGCTGT	CCGAAAACCT
	3051	TAATTTTACC	CTGCAAAACG	AACACGTCGA	TGCGGGCGCG	TGGCGTTACC
	3101	AACTCATCCG	CAAAGACGGC	GAGTTCGCGC	TGCATAATCC	GGTCAAAGAA
65	3151	CAAGAGCTTT	CCGACAAACT	CGGCAAGGCA	GAAGCCAAAA	AACAGGCGGA
	3201	AAAAGACAAC	GCGCAAAACC	TTGACGCGCT	GATTGCGGCG	GGGCGCGATG
	3251	CCGCCGAAAA	GACAGAAAGC	GTTGCCGAAC	CGGCCCGGCN	GGCAGGCGGG
	3301	GAAAATGTCT	GCATTATGCA	GGCGGAGGAA	GAGAAAAAAC	GGGTGCAGGC
	3351	GGATAAAGAC	AGCGCNTTGG	CGAAACAGCG	CGAAGCGGAA	ACCCGGCCGG
	3401	NTACCACCGC	CTTCCCCCGC	GCCCGCNGCG	CCCGCCGGGA	TTTGCCGCAA
70	3451	CCGCAGCCCC	AACCGCAACC	TCAACCCCAA	CCGCAGCGCG	ACCTGATNAG
	3501	CCGTATATGCC	AATAGCGGTT	TGAGTGAATT	TTCCGCCACG	CTCAACAGCG
	3551	TTTTCGCCGT	ACAGGACGAA	TTGGACCGCG	TGTTTGCCGA	AGACCGCCCG



-365-

5  
 10  
 15

```

3601 AACGCNGTTT GGACAAGCNG CATCCGGNAC ACCAAACACT ACCGTTTCGCA
3651 AGATTTCCTGC GCCTACCGCC AACAAACCGA CCTGCGCCAA ATCGGTATGC
3701 AGAAAAAACCCT CGGCAGCGGG CGCGTCGGCA TCCTGTTTTC GCACAACCGG
3751 ACCGAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA
3801 CGGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTTCGAC ATCGGCATCA
3851 GCACGGGCGC GGGTTTTAGC AGCGGCANTC TNTCAGACGG CATCGGAGGC
3901 AAAATCCGCC GCCGCGTGCT GCATTACGGC ATTCAGGCAC GATACCGCGC
3951 CGGTTTCGGC GGATTTCGGCA TCGAACCGTA CATCGGCGCA ACGCGCTATT
4001 TCGTCCAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC
4051 GGTCTTGCGT TCAACCGNTA CCGNGCGGGC ATTAAGGCAG ATTATTTCATT
4101 CAAACCGGCG CAACACATNT CCATCACNCC TTATTTNAGC CTGTCCTATA
4151 CCGATGCCGC TTCGGGCAAA GTCCGAACAC GCGTCAATAC CGCNGTATTG
4201 GCTCAGGATT TCGGCAAAAC CCGCAGTGC GAATGGGGCG TAAACGCCGA
4251 AATCAAAGGT TTCACGCTGT CCNTCCACGC TGCCGCCGCC AAAGGNCCGC
4301 AACTGGAAGC GCAACACAGC GCGGCATCA AATTAGGCTA CCGCTGGTAA
  
```

This encodes a protein having amino acid sequence <SEQ ID 652>:

20  
 25  
 30  
 35  
 40  
 45

```

1  MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
51  YQYYRDFAE NKGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVS RNG
101 VAALVG DQYI VSAHNGGYN NVDFGAEGXN PDQHRFSYQI VKRNNYKPDN
151 SHPYNGDXHM PRLHKFVTD AEPVEMTSDMR GNTYS DKEY PERVRIGSGH
201 HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVXSLSG DVRHANDYGP
251 MPIAGAAGDS GSPMFIYDKT NNKWLNLGV LQTGYPYSGRE NGFOLIRKDW
301 FYDDIYRGDT HTVXFEP RSN GHFSFTSNNN GTGTVTETNE KVS NPKLKVQ
351 TVRLFDESLN ETDKEPVYAA GGVNQYRPERL NNGENLSFID YGNGKLILSN
401 NINQGAGGLY FEGDFTVSPE NNETWQAGAV HISEDSTVTW KVNGVANDRL
451 SKIGKGT LHV QAKGENQGS I SVGDGTVILD QQADDKGGKQ AFSEIGLXSG
501 RGTVQLNADN QFNPDKLYFG FRGGRDLN G HSLSFHRIQN TDEGAMIXXH
551 NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR
601 LNLVYQPAE DRTXLLSGGT NLNGNITQTN GKLFSSGRPT PHAYNHLGSG
30 651 WSKMEGI PQG EIVWDNDWIX RTFKAENFHI QGGQAVISRN VAKVEGD XHL
701 SNHAQAVFGV APHQSH TICT RSDWTGLTNC VEXXITDDKV IASLTKTDXS
751 GXVXLXXXXX XXLXGXAXLX GNLSANGDTR YTVSHNATQN GNLSLVGNAQ
801 ATFNQATLNG NX SXSGNASF NLSNNAQNG SLTSLDNAKA NVSHSALNGN
851 VSLADKAVFH FENS RFTGQL SGSKXTALHL KDEWTLPSG TELGNLNDN
35 901 ATITLNSAYR HDAAGAQTGX VSDTPRRRSR RSLLSVTPPT SVESRFNTLT
951 VNGKLN XQGT FRFMSE LFGY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD
1001 QLTVEGKDN KPLSEN LNF T LQNEHVDAGA WRYQLIRKDG EFR LHNPFVKE
1051 QELSDKLGA EAKKQAEKDN AQLDALIAA GRDAAEKTES VAEFARXAGG
1101 ENVGIMQAE EKKRVQADKD SALAKQREAE TRPXTTAFPR ARKARRDL PQ
40 1151 PQPQPQPQ PQDLXSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR
1201 NAVWTSXIRX TKHYRSQDFR AYRQQTDLRQ IGMQKNLGSG RVGILFSHNR
1251 TENXFDDGIG NSARLAHGA V FCQYIGRFD IGISTGACFS SGXLSDGIGG
1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQKADY RYENVNIATP
1351 GLAFNRYRAG IKADYSFKPA QHXSITPYXS LSYTDAASGK VRTRVNTAVL
45 1401 AQDFGKTRSA EWGVNAEIKG FTLSXHAAAA KGPQLEAQHS AGIKLGYRW*
  
```

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

50  
 55  
 60  
 65

```

              10      20      30      40      50      60
orfla.pep    MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAE
              10      20      30      40      50      60
orfl1-1      MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAE

              70      80      90      100     110     120
orfla.pep    KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVG DQYIVSAHNGGYN
              70      80      90      100     110     120
orfl1-1      KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVG DQYIVSAHNGGYN

              130     140     150     160     170     179
orfla.pep    NVDFGAEGXNPDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTD AEPVEMTSDM
              130     140     150     160     170     180
orfl1-1      NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTD AEPVEMTSYM
  
```

-366-

5	orfla.pep	180	190	200	210	220	230
	orfl-1	190	200	210	220	230	240
10	orfla.pep	240	250	260	270	280	290
	orfl-1	250	260	270	280	290	
15	orfla.pep	300	310	320	330	340	350
	orfl-1	300	310	320	330	340	350
20	orfla.pep	360	370	380	390	400	410
	orfl-1	360	370	380	390	400	410
25	orfla.pep	420	430	440	450	460	470
	orfl-1	420	430	440	450	460	470
30	orfla.pep	480	490	500	510	520	530
	orfl-1	480	490	500	510	520	530
35	orfla.pep	540	550	560	570	580	590
	orfl-1	540	550	560	570	580	590
40	orfla.pep	600	610	620	630	640	650
	orfl-1	600	610	620	630	640	650
45	orfla.pep	660	670	680	690	700	710
	orfl-1	660	670	680	690	700	710
50	orfla.pep	720	730	740	750	760	770
	orfl-1	720	730	740	750	760	770
55	orfla.pep	780	790	800	810	820	830
	orfl-1	780	790	800	810	820	830
60	orfla.pep	840	850	860	870	880	890
	orfl-1	840	850	860	870	880	890
65	orfla.pep	900	910	920	930	940	950
	orfl-1	900	910	920	930	940	950
70	orfla.pep	960	970	980	990	1000	1010
	orfl-1	960	970	980	990	1000	1010

-367-

		900	910	920	930	940
	orf1a.pep	TELGNLNLDNATITLNSAYRHDAAAGATGXVSDTPRRRSRRS---	LLSVTPPTSVESRFN			
5	orf1-1	TELGNLNLDNATITLNSAYRHDAAAGATGSATDAPRRRSRRSRRSLLSVTPPTSVESRFN				
		900	910	920	930	940
	orf1a.pep	950	960	970	980	990
		TLTVNGKLNQQTFRFMSELFGYRSDKLKLAESSEGYTLAVNNTGNEPVS	LDQLTVVEG			
10	orf1-1	TLTVNGKLNQQTFRFMSELFGYRSDKLKLAESSEGYTLAVNNTGNEPASLE	QQLTVVEG			
		960	970	980	990	1000
	orf1a.pep	1010	1020	1030	1040	1050
		KDNKPLSENLFNLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKL	GKAEAKKQAE			
15	orf1-1	KDNKPLSENLFNLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKL	GKAEAKKQAE			
		1020	1030	1040	1050	1060
	orf1a.pep	1070	1080	1090	1100	1110
		KDNAQSLDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAE	EEKKRVQADKDSALAKQR			
20	orf1-1	KDNAQSLDALIAAGRDAVEKTESVAEPARQAGGENVGIMQAE	EEKKRVQADKDTALAKQR			
		1080	1090	1100	1110	1120
	orf1a.pep	1130	1140	1150	1160	1170
		EAETREXTTAFPRARXARRDLPPQPQPQPQQRDLXSRYAN	SGLSEF	SATLNSVFAV		
25	orf1-1	EAETREPTTAFPRARRARRDLPPQLQPQPQPQ--QRDLISRYAN	SGLSEF	SATLNSVFAV		
		1140	1150	1160	1170	1180
	orf1a.pep	1190	1200	1210	1220	1230
		QDELDRVFAEDRRNAVWTSXIRXTHYRSQDFRAYRQQTDLRQIGM	QKNLGSGR	VGILFS		
35	orf1-1	QDELDRVFAEDRRNAVWTSXIRXTHYRSQDFRAYRQQTDLRQIGM	QKNLGSGR	VGILFS		
		1200	1210	1220	1230	1240
	orf1a.pep	1250	1260	1270	1280	1290
		HNRTENXFDDGIGNSARLAHGAVFGQYIGRFDIGISTGAGFSSG	XLSDGIGG	KIRRRVL		
40	orf1-1	HNRTENTFDDGIGNSARLAHGAVFGQYIGRFDIGISTGAGFSSG	XLSDGIGG	KIRRRVL		
		1260	1270	1280	1290	1300
	orf1a.pep	1310	1320	1330	1340	1350
		HYGIQARYRAGFGGFGIEPHYGATRYFVQKADYRYENVNIATPGLAF	NRYRAGIKADYSF			
45	orf1-1	HYGIQARYRAGFGGFGIEPHYGATRYFVQKADYRYENVNIATPGLAF	NRYRAGIKADYSF			
		1320	1330	1340	1350	1360
	orf1a.pep	1370	1380	1390	1400	1410
		KPAQHXSITPYXSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEW	GVNAEIKGFTLSXHA			
50	orf1-1	KPAQHISITPYLSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEW	GVNAEIKGFTLSLHA			
		1380	1390	1400	1410	1420
	orf1a.pep	1430	1440	1450		
		AAAKGPQLEAQHSAGIKLGYRWX				
55	orf1-1	AAAKGPQLEAQHSAGIKLGYRWX				
		1440	1450			
60						

Homology with adhesion and penetration protein hap precursor of *H.influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

orf1	23	FXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDF	FAENKKGKFAVGAKDIEVYNKKGELVG	82
		F +L C+S GI QAWAGHTYFGI+YQYYRDF	FAENKKGK VGAK+IEVYNK+G+LVG	
hap	6	FRLNFLTACVSLGIASQAWAGHTYFGIDYQYYRDF	FAENKKGKFTVGAKNIEVYNKEGQLVG	65
orf1	83	KSMTKAPMIDFSVVSRRNGVAALVGVQYIVSVAHNGG	YNNVDFGAEGXNIXDQXRXTYKIV	142
		SMTKAPMIDFSVVSRRNGVAALVG QYIVSVAHNGG	YN+VDFGAEG N DQ R TY+IV	
hap	66	TSMTKAPMIDFSVVSRRNGVAALVGDQYIVSVAHNGG	YNDVDFGAEGRN-PDQHRFTYQIV	124

5	orf1	143	KRNNYKAGTKGHPYGGDYHMPRLHKXVTDAPVEMTSYMDGRKYIDQNNYPDRVRIGAGR	202
	hap	125	KRNNY+A + HPY GDYHMPRLHK VT+AEPV MT+ MDG+ Y D+ NYP+RVRIG+GR	184
10	orf1	203	QYWRSDDEPNNRESSYHIA-----	222
	hap	185	QYWR+D+DE N SSY+++	244
15	orf1	223	-----SGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVF	277
	hap	245	SGSPMFIYDA+K++WLIN VLQTG+P+ G+ NGFQL+R++WFY+E+ A DT SVF	304
20	orf1	278	--YEPQNGKYSFNDDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHA	334
	hap	305	Y P NG YSF +N+GTGK+ + + + + TV+LFN SL++TA+E V A	363
25	orf1	335	AGGVNSYRPRLNNGENISFIDEKGELILTSNINQGAGGLYFQGDFTV-SPENNETWQGA	393
	hap	364	A G N Y+PR+ G+NI D+GK L + +NINQGAGGLYF+G+F V +NN TWQGA	423
30	orf1	394	GVHISEDSTVTWKVNGVANDRLSKIGKGT	423
	hap	424	GV I +D+TV WKV+ NDRLSKIG GTL	453
Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:				
35	Orf1	41	DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASFNLSDHAVQNGSLTLS	98
	hap	733	DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS	792
40	orf1	99	GNAKANVSHSALNGNVSLADKAVFHESSRFTGQISGGKDTALHLKDSEWTLPSGXELGN	158
	hap	793	+A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N	852
45	orf1	159	LNLDNATITLNSAYRHDAAGAQTGSATDAPXXXXXXXXXXLLXVTPPTSVESRENTLTVN	218
	hap	853	L L+N+T+TLNSAY + S+ +AP L T PTS E RENTLTVN	899
50	orf1	219	GKLNQGTFRFMSELFYGRSDKLKLAESSEGYTLAVNNTGNEPASLEQLTVVEGKDNKP	278
	hap	900	GKL+GQGTFF+ F S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP	959
55	orf1	279	LSENLNFTLQNEHVDAGA	296
	hap	960	LS+ L FTL+N+HVDAGA	977
Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:				
60	Orf1	1	LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGILFSHNR	60
	hap	1135	LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R	1194
65	orf1	61	TENTFDDGIGNSARLAHGAVFQGYGIDRFYXXXXXXXXXXXXXXXXXIGKXRRRVLHYG	120
	hap	1195	++NTFD+ + N A L + F QY K R+ ++YG	1254
70	orf1	121	IQARYRAGFGGEGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA	180
	hap	1255	+ A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P	1314
75	orf1	181	QHISITPYLSLSYTDASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAA	240
	hap	1315	+IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + +	1374
80	orf1	241	KGPQLEAQSAGIKLGYRW	259
	hap	1375	+G QL Q + G+KLGyrw	1393

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

5	orf1.pep	MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN	60
	orf1ng	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN	60
10	orf1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYN	120
	orf1ng	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALAGDQYIVSVAHNGGYN	120
15	orf1.pep	NVDGFAEGXNIXDQXRTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTDAPVEMTSY	180
	orf1ng	NVDGFAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAPVEMTSY	179
20	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNRESSYHIAS-----	223
	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWRSDDEPNRESSYHIASAYSWLVGGNTFAQNGSG	239
25	orf1.pep	-----GSPMFIYDAQKQKWLINGVLQTGNPYIGKSNG	255
	orf1ng	GGTVNLGSEKIKHSPYGFLLPTGGSGSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNG	289
30	orf1.pep	FQLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT	315
	orf1ng	FQLVRKDWFYDEIFAGDTHSVFYEPHONGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRT	359
35	orf1.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQGAGGLY	375
	orf1ng	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDKGGELILTSNINQGAGGLY	375
40	orf1.pep	FQGDFTVSPENNETWQAGAGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
	orf1ng	FEGNFTVSPKNNETWQAGAGVHISDGSSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
45	orf1.pep	// DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
	orf1ng	FGVAPHQSHTICTRSDWTGLTSCTEKTITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774
50	orf1.pep	TLNGNLSANGDTR-YTVSHNATQNGNXSLVXNAQATFNQATLNGNTSASGNASFNLSDBA	803
	orf1ng	TFNGNL-VQAETRTIRLANATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNA	833
55	orf1.pep	VQNGSLTSLGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWT	863
	orf1ng	VQNGSLTSLSDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWT	893
60	orf1.pep	LPSGXELGNLNDNATITLNSAYRHAAGAQTGSATDAPRRRSRRSRLXVTPPTSVE	923
	orf1ng	LPSGTELGNLNDNATITLNSAYRHAAGAQTGSAADAPRRRSRRS---LLSVTPPTSVE	950
65	orf1.pep	SRFNTLTVNGKLNQGTFRFMSELFYRSCLKLAESSEGYTLAVNNTGNEPASLEQLT	983
	orf1ng	SRFNTLTVNGKLNQGTFRFMSELFYRSCLKLAESSEGYTLAVNNTGNEFVSLEQLT	1010
70	orf1.pep	VVEGKDNKPLSENINFTLQNEHVDAGAW	1011
	orf1ng	VVEGKDNTPLSENINFTLQNEHVDAGAWRYQLIRKDGFRHLNPNVKEQELSDKLKGAGET	1070
75	orf1.pep	// LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1211
	orf1ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1239
80	orf1.pep	AYRQQTDLRQIGMQKNLGSRGVILFSHNRTENTFDDGIGNSARLAHGAVFQYQIDRFY	1271
	orf1ng	AYRQQTDLRQIGMQKNLGSRGVILFSHNRTGNTFDDGIGNSARLAHGAVFQYQIGRFD	1299

	orfl.pep	IGISAGAGFSSGSLSDGIGXKXRRRLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1331
	orfl.ng	IGISAGAGFSSGSLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1359
5	orfl.pep	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAASKVTRVNTAVL 1391
	orfl.ng	RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAASKVTRVNTAVL 1419
10	orfl.pep	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGPQLEAQSAGIKLGYRW 1440
	orfl.ng	AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGPQLEAQSAGIKLGYRW 1468

The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCTAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCCGCTTA	CTTAGCCATA	TGCCTGTCTG
15	101	TCGGCATTCT	GCCCCAAGCC	CGGGCGGGAC	ACACTTATTT	CGGCATCAAC
	151	TACCAATACT	ATCGCGACTT	TGCCGAAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGGTTT	ACAACAAAAA	AGGGGAGTTG	GTCGGCAAAT
	251	CGATGACGAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTATC	GCGTAACGGC
20	301	GTGGCGGCAT	TGGCGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
	351	CGGCTATAAC	AATGTTGATT	TTGGTTCGGA	GGGAAGCAAT	CCCGATCAGC
	401	ACCGCTTTTC	TTACCAAATT	GTGAAAAGAA	ATAATTATAA	AGCAGGGACT
	451	AACGGCCATC	CTTATGGCGG	CGATTATCAT	ATGCCCGGTT	TGCACAAATT
	501	TGTCACAGAT	GCAGAACCTG	TTGAGATGAC	CAGTTATATG	GATGGGTGGA
25	551	AATACGCTGA	TTTAAATAAA	TACCCGTATC	GTGTTTCAAT	CGGAGCAGGC
	601	AGACAATATT	GGCGGTCTGA	TGAAGACGAA	CCCAATAACC	GCGAAAGTTC
	651	ATATCATATT	GCAAGCGCAT	ATTCTTGGCT	CGTCGGTGGC	AATACCTTTG
	701	CACAAAATGG	ATCAGGTGGT	GGCACAGTCA	ACTTAGGTAG	CGAAAAAATT
	751	AAACATAGCC	CATATGGTTT	TTTACCAACA	GGAGGCTCAT	TTGGCGACAG
	801	TGGCTCACCA	ATGTTTATCT	ATGATGCCCA	AAAGCAAAAG	TGGTTAATTA
30	851	ATGGGGTATT	GCAACAGGGC	AACCCCTATA	TAGGAAAAAG	CAATGGCTTC
	901	CAGCTAGTTC	GTAAAGATTG	GTTCTATGAT	GAAATCTTTG	CTGGAGATAC
	951	CCATTACAGT	TTCTACGAAC	CACATCAAAA	TGGGAAATAC	TTTTTTAACG
	1001	ACAATAATAA	TGGCGCAGGA	AAAATCGATG	CCAAACATAA	ACACTATTCT
	1051	CTACCTTTATA	GATTAAAAAC	ACGAACCGTT	CAATTGTTTA	ATGTTTCTTT
35	1101	ATCCGAGACA	GCAAGAGAAC	CTGTTTATCA	TGCTGCAGGT	GGGGTCAACA
	1151	GTTATCGACC	CAGACTGAAT	AATGGAGAAA	ATATTTCTCT	TATTGACAAA
	1201	GGAAAAGGTG	AATTGATACT	TACCAGCAAC	ATCAACCAAG	GCGCGGGCGG
	1251	TTTGTATTTT	GAGGGTAATT	TTACGGTCTC	GCCTAAAAAC	AACGAAACCT
	1301	GGCAAGGCGC	GGGCGTTCAT	ATCAGTGATG	GCAGTACCGT	TACTTGGAAG
40	1351	GTAAACGGCG	TGGCAAACGA	CCGCCTGTCC	AAAATCGGCA	AAGGCACGCT
	1401	GCTGGTTCAA	GCCAAAGGGG	AAAACCAAGG	CTCGGTCAGC	GTGGGCGACG
	1451	GTAAAGTCAT	CTTAGATCAG	CAGGCGGACG	ATCAAGGCAA	AAAACAAGCC
	1501	TTTAGTGAAA	TCGGCTTGGT	CAGCGGCAGG	GGGACGGTGC	AACTGAATGC
	1551	CGATAATCAG	TTCAACCCCG	ACAACTCTTA	TTTCGGCTTT	CGCGGCGGAC
45	1601	GTTTGGATTT	GAACGGGCAT	TCGCTTTCGT	TCCACCGCAT	TCAAAATACC
	1651	GATGAAGGGG	CGATGATTGT	CAACCACAAT	CAAGACAAAG	AATCCACCGT
	1701	TACCATTACA	GGCAATAAAG	ATATTACTAC	AACCGGCAAT	AACAACAAC
	1751	TGGATAGCAA	AAAAGAAATT	GCCTACAACG	GTGTTTGGG	CGAGAAAGAT
	1801	GCAACCAAAA	CGAACGGGCG	GCTCAATCTG	AATTACCAAC	CGGAAGAAGC
50	1851	GGATCGCACT	TTACTGCTTT	CCGGCGGAAC	AAATTAAAC	GGCAATATCA
	1901	CGCAAAACAA	CGGCAAACTG	TTTTTTCAGC	GCAGACCGAC	ACCGCACGCC
	1951	TACAATCATT	TAGGAAGCGG	GTGGTCAAAA	ATGGAAGGTA	TCCCACAAGG
	2001	AGAAATCGTG	TGGGACAACG	ATTGGATCGA	CCGCACATTT	AAAGCGGAAA
	2051	ACTTCCATAT	TCAGGGCGGA	CAAGCGGTGG	TTTCCCGCAA	TGTTGCCAAA
55	2101	GTGGAAGGCG	ATTGGCATTT	AAGCAATCAC	GCCCAAGCAG	TTTTCGGTGT
	2151	CGCACCGCAT	CAAAGCCACA	CAATCTGTAC	ACGTTTCGGC	TGGACGGGTC
	2201	TGACAAGTTG	TACCGAAAAA	ACCATTACCG	ACGATAAAGT	GATTGCTTCA
	2251	TTGAGCAAGA	CCGACATCAG	AGGCAATGTC	AGCCTTGCCG	ATCACGCTCA
	2301	TTTAAATCTC	ACAGGACTTG	CCCACTCAA	CGGCAATCTT	AGTGACGGCG
60	2351	GAGACACGCA	CTATACGGTT	ACGCGCAACG	CCACCAAAA	CGGCAACCTC
	2401	AGCCTCGTGG	GCAATGCCCA	AGCAACATTT	AATCAAGCCA	CATTAAACGG
	2451	TAACACATCG	GCTTCGGACA	ATGCTTCATT	TAATCTAAGC	AACAACGCCG
	2501	TACAAAACGG	CAGTCTGACG	CTTTCGGACA	ACGCTAAGGC	AAACGTAAGC
	2551	CATTCCGCAC	TCAACGGCAA	TGTCTCCCTA	GCCGATAAGG	CAGTATTCCA
65	2601	TTTTGAAAAC	AGCCGCTTTA	CCGGAAAAAT	CAGCGGCGGC	AAGGATACGG
	2651	CATTACACTT	AAAAGACAGC	GAATGGACGC	TGCCGTCGGG	CACGGAATTA
	2701	GGCAATTTAA	ACCTTGACAA	CGCCACCATT	ACACTCAATT	CCGCTATCG
	2751	ACACGATGCG	GCAGGCGCGC	AAACCGGCAG	TGCGGCAGAT	GCGCCGCGCC
	2801	GCCGTTCCGC	CCGTTCCCTA	TTATCCGTTA	CGCCGCCAAC	TTCCGCAGAA
70	2851	TCCCGTTTCA	ACACGCTGAC	GGTAAACGGC	AAATTGAACG	GTCAGGGAAC

-371-

2901 ATTCGCTTT ATGTCGGAAC TCTTCGGCTA CCGCAGCGGC AAATTGAAGC  
 2951 TGGCGGAAAG TTCCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCGGC  
 3001 AACGAACCCG TAAGTCTCGA GCAATTGACG GTAGTGGAAG GAAAAGACAA  
 3051 CACACCGCTG TCCGAAAATC TTAATTTTAC CCTGCaaaAc gaacacgtcg  
 5 3101 atgccggcgc atggCGTTAT CAGCTTATCC gcaaagacgG CGAGTTCcGc  
 3151 CTGCATAATC CGGTCAAAGA ACAAGAGCTT TCCGACAAAC TCGGCAAGgc  
 3201 gggagaaACA GAgggccgcT TGACGGCAAA ACAGGCacaA CTTGCCGCCA  
 3251 AAcaacaggc ggaaaAAGAC AACgcgcaaa gccttgAcgc gctgattgcg  
 3301 gCggggcgca atgccaccga AAAGGCAGaa agtgttgccg aaccgGCCCG  
 10 3351 GCAGGCAGGC GGGGAAAAtg ccgGCATTAT GCAGGCGGAG GAAGAGAAAA  
 3401 AACGGGTGCA GCGGATAAA GACACCGCCT TGGCGAAACA GCGCGAAGCG  
 3451 GAAACCCGGC CGGCTACCAC CGCCTTCCCC CGCGCCCGCC GCGCCCGCCG  
 3501 GGATTTCGCG CAACCGCAGC CCCAACCGCA ACCCCAACCG CAGCGCGACC  
 3551 TGATCAGCCG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGTC  
 15 3601 AACAGCGTTT TCGCCGTACA GGACGAATTG GACCGCGTGT TTGCCGAAGA  
 3651 CCGCCGCAAC GCCGTTTGA CAAGCGGCAT CCGGGACACC AAACACTACC  
 3701 GTTCGCAAGA TTTCCGCGCC TACCGCCAAC AAACCGACCT GCGCCAAATC  
 3751 GGTATGCAGA AAAACCTCGG CAGCGGGCGC GTCGGCATCC TGTTTTCGCA  
 3801 CAACCGGACC GGAAACACCT TCGACGACGG CATCGGCAAC TCGGCACGGC  
 20 3851 TTGCCACCGG TGCCGTTTTC GGGCAATACG GCATCGGCAG GTTCGACATC  
 3901 GGCATCAGCG CGGGCGCGGG TTTTAGTAGC GGCAGCCTTT CAGACGGCAT  
 3951 CAGAGGCAAA ATCCGCGGCC GCGTGCTGCA TTACGGCATT CAGGCAAGAT  
 4001 ACCGCGCAGG TTTCCGCGGA TTCGGCATCG AACCGCACAT CGGCGCAACG  
 4051 CGCTATTTTC TCCAAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC  
 25 4101 CACCCCGGGC CTTGCAATCA ACCGCTACCG CGCGGGCATT AAGGCAGATT  
 4151 ATTCATTCAA ACCGGCGCAA CACATTTCCA TCACGCCTTA TTTGAGCCTG  
 4201 TCCTATACCG ATGCGGCTTC CGGCAAAAGT CGAACGCGCG TCAATACCGC  
 4251 CGTATTGGCG CAGGATTTCG GCAAAACCCG CAGTGCGGAA TGGGGCGTAA  
 4301 ACGCCGAAAT CAAAGGTTTC ACGCTGTCCC TCCACGCTGC CGCCGCCAAG  
 30 4351 GGGCCGCAAT TGAAGCGCA GCACAGCGCG GGCATCAAAT TAGGCTACCG  
 4401 CTGGTAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

1 MKTTDKRTTE THRKAPKTGR IREFSPAYLAI CLSFGILPQA RAGHTYFGIN  
 51 YQYYRDFAEN KGFVAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVRNG  
 101 VAALAGDQYI VSVAHNGGYN NVDFGAEGSN PDQHRFSYQI VKRNNYKAGT  
 151 NGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGWKYADLNK YPDRVRIGAG  
 201 RQYWRSDDEE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI  
 251 KHSPYGFLEPT GGSFGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNGF  
 301 QLVRKDWFYD EIFAGDTHSV FYEPHQNGKY FFNDNNNGAG KIDAKHKHYS  
 40 351 LPYRLKTRTV QLFNVSLSET AREPVYHAAG GVNYSRPRLN NGENISFIDK  
 401 KGCELILTSN INQAGGLYF EGNFTVSPKN NETWQAGVH ISDGSVTWTK  
 451 VNGVANDRLS KIGKGTLLVQ ARGENQGSVS VGDGKVILDQ QADDQKKQA  
 501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT  
 551 DEGAMIVNHN QDKESTVTIT GNKDITTTGN NNNLDSKKEI AYNGWFGEKD  
 601 ATKTNGLNL NYPPEEADRT LLLSGGTNLN GNITQTNGKL FFSGRPTFHA  
 651 YNHLGSGWSK MEGIPQGEIV WDNDWIDRTF KAENFHIQGG QAVVSRNVAK  
 701 VEGDWHLNSH AQAVFCVAPH QSHTICTRSD WTGLTSCTEK TITDDKVIAS  
 751 LSKTDVRGNV SLADHAHLNL TGLATFNGNL VQAETrTIRL RANATQNGNL  
 801 SLVGNAQATF NQATLNGNTS ASDNASFNLS NNAVQNGSLT LSDNAKANVS  
 50 851 HSALNGNVS ADKAVPHFEN SRFTGKISGG KDTALHLKDS EWTLPSTGTEL  
 901 GNLNLDNATI TLNSAYRHDA AGAQTGSAAD APRRRSRRLS LSVTPPTSAE  
 951 SRFNTLTVNG KLNGQGTFRF MSELFGYRSG KLKLAESSEG TYTLAVNNTG  
 1001 NEPVSLQLT VVEGKDNTPL SENLNFTLQN EHVDAGAWRY QLIRKDGEFR  
 1051 LHPVKEQEL SDKLGKAGET EAALTAKQAO LAAKQQAED NAQSLDALIA  
 55 1101 AGRNATEKAE SVAEPARQAG GENAGIMQAE EEKRVQADK DTALAKQREA  
 1151 ETRPATTAFF RARRARRDLP QPQPQPQPQ QRDLSRYAN SGLSEFSATL  
 1201 NSVFAVQDEL DRVFAEDRRN AVWTSGIRDY KHYRSQDFRA YRQQTDLRQI  
 1251 GMQKNLGSGR VGILFSHNRT GNTFDDGIGN SARLAHGAVF GQYIGREFDI  
 1301 GISAGAGFSS GSLSDGIRGK IRRRVLYHYG QARYRAGFGG FGIEPHIGAT  
 60 1351 RYFVQKADYR YENVNIATPG LAFNRYRAGI KADYSEKPAQ HISITPYLSL  
 1401 SYTDAASGKV RTRVNTAVLA QDFGKTRSAE WGVNAEIKGF TSLSLHAAAAA  
 1451 GPQLEAQHSA GIKLGYRW\*

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

65 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

-372-

		10	20	30	40	50	60
	orf1-1.pep	MKT	TDK	RTE	THRK	APKT	GRIR
	orf1ng-1	MKT	TDK	RTE	THRK	APKT	GRIR
5		10	20	30	40	50	60
	orf1-1.pep	70	80	90	100	110	120
	orf1ng-1	70	80	90	100	110	120
10		130	140	150	160	170	180
	orf1-1.pep	NVD	FGE	GRNP	DQHR	FYTK	IVKR
	orf1ng-1	NVD	FGE	GRNP	DQHR	FYTK	IVKR
15		130	140	150	160	170	180
	orf1-1.pep	190	200	210	220	230	240
	orf1ng-1	190	200	210	220	230	240
20		250	260	270	280	290	300
	orf1-1.pep	GTV	NLG	SEK	IKHS	PYGF	LPTG
	orf1ng-1	GTV	NLG	SEK	IKHS	PYGF	LPTG
25		250	260	270	280	290	300
	orf1-1.pep	310	320	330	340	350	360
	orf1ng-1	310	320	330	340	350	360
30		370	380	390	400	410	420
	orf1-1.pep	QLF	NVSL	SETA	REP	VYHA	AGGV
	orf1ng-1	QLF	NVSL	SETA	REP	VYHA	AGGV
35		370	380	390	400	410	420
	orf1-1.pep	430	440	450	460	470	480
	orf1ng-1	430	440	450	460	470	480
40		490	500	510	520	530	540
	orf1-1.pep	VGD	GT	VILD	QQAD	DKGK	QAFS
	orf1ng-1	VGD	GT	VILD	QQAD	DKGK	QAFS
45		490	500	510	520	530	540
	orf1-1.pep	550	560	570	580	590	600
	orf1ng-1	550	560	570	580	590	600
50		610	620	630	640	650	660
	orf1-1.pep	TTK	TNG	RNL	NLY	QPAE	DRTL
	orf1ng-1	TTK	TNG	RNL	NLY	QPAE	DRTL
55		670	680	690	700	710	720
	orf1-1.pep	KEG	IP	QGE	IV	WDND	WINR
	orf1ng-1	KEG	IP	QGE	IV	WDND	WINR
60		670	680	690	700	710	720
	orf1-1.pep	KEG	IP	QGE	IV	WDND	WINR
	orf1ng-1	KEG	IP	QGE	IV	WDND	WINR
65		670	680	690	700	710	720
	orf1-1.pep	KEG	IP	QGE	IV	WDND	WINR
	orf1ng-1	KEG	IP	QGE	IV	WDND	WINR
70		670	680	690	700	710	720
	orf1-1.pep	KEG	IP	QGE	IV	WDND	WINR
	orf1ng-1	KEG	IP	QGE	IV	WDND	WINR



-373-

		730	740	750	760	770	780
	orf1-1.pep	QSHTICTRSDWTGLTNCVEKTIITDDKVIASLT	TKTDISGNVDLADHAHLNLTGLATLNGNL				
5	orf1ng-1	QSHTICTRSDWTGLTSCTEKTIITDDKVIASL	SKTDIRGNVSLADHAHLNLTGLATLNGNL				
		730	740	750	760	770	780
	orf1-1.pep	790	800	810	820	830	840
10	orf1ng-1	SANGDTRYTVSHNATQNGNLSLVGNAQATFNQAT	LNNGNTSASGNASFNLS	SDHAVQNGSLT			
		790	800	810	820	830	840
	orf1-1.pep	850	860	870	880	890	900
15	orf1ng-1	LSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTEL					
		850	860	870	880	890	900
	orf1-1.pep	910	920	930	940	950	960
20	orf1ng-1	GNLNLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRRSLLSVTPPTSVESRFTLT					
		910	920	930	940	950	
25	orf1-1.pep	970	980	990	1000	1010	1020
	orf1ng-1	VNGKLNQGTFRFMSELFGYRSDKLKLAESSEGTYTLAVNNTGNEPASLEQLTVVEGKDN					
30		960	970	980	990	1000	1010
	orf1-1.pep	1030	1040	1050	1060	1070	
35	orf1ng-1	KPLSENLFNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKA-----					
		1020	1030	1040	1050	1060	1070
	orf1-1.pep	1080	1090	1100	1110	1120	
40	orf1ng-1	----EAKKQAEKDNAQSLDALIAAGRDAVEKTESVAEPARQAGGENVGIMQAEKKRVQ					
		1080	1090	1100	1110	1120	1130
	orf1-1.pep	1130	1140	1150	1160	1170	1180
45	orf1ng-1	ADKDTALAKQREAETRPATTAFFPRARRARDLPQLQPQPQPQORDLISRYANGLSEFS					
		1140	1150	1160	1170	1180	1190
	orf1-1.pep	1190	1200	1210	1220	1230	1240
50	orf1ng-1	ATLNSVFAVQDELDRVFAEDRRNAVWTS	SGIRD	TKHYRSQDFRAYRQQTDLRQIGMQKNLG			
		1200	1210	1220	1230	1240	1250
55	orf1-1.pep	1250	1260	1270	1280	1290	1300
	orf1ng-1	SGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGI					
60		1260	1270	1280	1290	1300	1310
	orf1-1.pep	1310	1320	1330	1340	1350	1360
65	orf1ng-1	GGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADRYENVNIATPGLAFNRYR					
		1320	1330	1340	1350	1360	1370
	orf1-1.pep	1370	1380	1390	1400	1410	1420
70	orf1ng-1	AGIKADYSFKPAQHISITPYLSLSYTDAA	SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI				
		1380	1390	1400	1410	1420	1430

```

      1430      1440      1450
orf1-1.pep  KGFTLSLHAAAAGKPQLEAQHSAGIKLGYRWX
             |||||
orf1ng-1    KGFTLSLHAAAAGKPQLEAQHSAGIKLGYRWX
      1440      1450      1460

```

10

SCORES Init1: 1104 Initn: 4632 Opt: 2680  
Smith-Waterman score: 5165; 55.7% identity in 1455 aa overlap

15

```

      10      20      30      40      50      60
orf1ng-1.pep MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYYRDAFEN
p45387          MKKTVFRLNFLTACISLGLIVSQAWAGHTYFGIDYQYYRDAFEN
      10      20      30      40

```

20

```

              70          80          90         100         110         120
orf1ng-1.pep KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSVSRNGVAALAGDQYIVSVAHNGGYN
|||||::||::||::||| ||||| ||||| ||||| : ||||| |||
p45387        KGKFTVGAQNIKVYNKGQLVGTSMTKAPMIDFSVSVSRNGVAALVENQYIVSVAHNVTGYT
              50          60          70          80          90         100

```

25

```

                                130      140      150      160      170      180
orf1ng-1.pep    NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTD AEPVEMTSYM
                :| | | | | |:| | | | |:| | | | |:| | | | |:| | | | |:| | | |
p45387          DVDFGAEGNPNPDQHRFTYKIVKRNNYKKD-NLHPYEDDYHNPR LHKFVTEAAPIDMTSNM
                110      120      130      140      150      160

```

30

```

                                190      200      210      220      230      240
orf1ng-1.pep  DGWKYADLNKYPDRVRIGAGROYWRSDEDEPNNRESSYHIASAYSWLVGGNFTFAQNGSGG
               :|  |  :||:||||:||||:|:|:|:  :  :||:|  :||:|  |  |  :|
p45387        NGSTYSDRTKYPERVRIGSGRQFWNRNDQKGD-----QVAGAYHYLTAGNTHNQRGAGN
               170      180      190      200      210

```

35

```

                250      260      270      280      290      300
orf1ng-1.pep  GTVNLGSEKIKHSPYGF LPTGGSGFDGSGSPMFIYDAQKQKWLINGVLQTGNPFYIGKSNGF
               |  ||:  | : || || :|| ||||| ||||| :||| ||| :| ||: || ||
p45387        GYSYLGGDVRKAGEYGPLPIAGSKGDSGSPMFIYDAEKQKWLINCILREGNPFEGKENG
               220      230      240      250      260      270

```

40

```

              310      320      330      340      350      360
orf1ng-1.pep QLVRKDWFYDEIFAGDTHSVFYEPHONGKYFFNDNNNGAGKIDAKHHYSLPYRLKTRTV
||| |::| ||| | : :| || : : |::| ::| :| ::| :
p45387        QLVRKSYP-DEIFERDLHTSLYTRAGNGVYTISGNDNGQGSITQKS---GIPSEIK---I
          280       290       300       310       320

```

45

```

                                370      380      390      400      410      419
orf1ng-1.pep  QLFNVSLSSETAREPVYHAA-GGVNSYRPLNNGENISFIDKKGKELILTSNINQGAGGLY
                | :||      :: :||      | | | |||||:: :| :| :| :||| |||
p45387        TLANMSLPLKEKDKVHNPRYDGPNIYSPRLNNGETLYFMDQKQGLSIFASDINQGAGGLY
              330      340      350      360      370      380

```

50

```

              420      430       440        450         460          470           479
orf1ng-1.pep FEGNFTVSPKNNETWQGAGVHISDGSVTTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV
|||||||:::||:|||||:::||:|||||:::||:|||||:::||:|||||:::||:|||||:::||:
p45387       FEGNFTVSPNSNQTWQGAGIHVSENSTVTWKVNGVEHDRLSKIGKGTLHVQAKGENKGI
              390      400       410        420         430          440

```

55

```

480      490      500      510      520      530      539
orf1ng-1.pep  SVGDGKVIILDQQADDQGKKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRDLNG
               |||||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
p45387        SVGDGKVILEQQADDQGNKQAFSEIGLVSGRGTVQLNDDKQFDTDKIFYFGFRGGRDLNG
450      460      470      480      490      500

```

60

[illegible]

65

70

-375-

5	orf1ng-1.pep	600 610 620 630 640 650	EKDATKTNGRLNLNYQPEEADRTLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSG
	p45387	570 580 590 600 610 620	ETDKNKHNGRLNLIYKPTTEDRTLLSGGTNLKGDITQTKGKLFFSGRPTPHAYNHLNKR
10	orf1ng-1.pep	660 670 680 690 700 710	WSKMEGIPQGEIVWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVEGDWHLNHAQAVFGV
	p45387	630 640 650 660 670 680	WSEMEGIPQGEIVWDHDWINRTFKAENFQIKGGS AVVSRNVSSIEGNWTVSNANATFGV
15	orf1ng-1.pep	720 730 740 750 760 770	APHQSHICTRSDWTGLTSCTEKTIITDDKVIASLSKTDIRGNVSLADHAHLNLTGLATLN
	p45387	690 700 710 720 730 740	VPNQNTICTRSDWTGLTTCQKVDLTDTKVINSIPKTQINGSINLTDNATANVKGLAKLN
20	orf1ng-1.pep	780 790 800 810 820 830	GNLSAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLNNAVQNG
	p45387	750 760 770	GNVTL-----TNHSQFTLSNNAQTIG
25	orf1ng-1.pep	840 850 860 870 880 890	SLTSLDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWTLPSG
	p45387	780 790 800 810 820 830	NIRLSDNSTATVDNANLNGNVHLTDSAQFSLKNSHFSHQIQGDKGTTVTLENATWTMPSD
30	orf1ng-1.pep	900 910 920 930 940 950	TELGNNLNDNATITLNSAYRHDAAQAQTGSAADAPRRRSRRSLLSVTPPTSASERFNTLT
	p45387	840 850 860 870	TTLQNLTNNSTITLNSAY-----SASSNNTPRRS---LETETTP TSAEHRFNTLT
40	orf1ng-1.pep	960 970 980 990 1000 1010	VNGKLNQGTFRFMSELFYRSGKLKLAESSEGTYTLAVNNTGNEPVSLEQLTVVEGKDN
	p45387	880 890 900 910 920 930	VNGKLSGQGTFTQFTSSLFGYKSDKLKLSNDAEGDYILSVRNTGKEPETLEQLTLVESKDN
45	orf1ng-1.pep	1020 1030 1040 1050 1060 1070	TPLSENLFNFTLQNEHVDAGAWRYQLIRKDGFEFLHNPVKEQELSDKLKGAGETEAALTAK
	p45387	940 950 960 970 980 990	QPLSDKLKFTLENDHVDAGALRYKLVKNDGFEFLHNPKEQELHNDLVRAEQAERTLEAK
50	orf1ng-1.pep	1080 1090 1100 1110 1120 1130	QAQLAAKQQAEDNAQSLDALIAAGR NAT-EKAESVAEPARQAGGENAGIMQAEEEKRV
	p45387	1000 1010 1020 1030 1040 1050	QVEPTAKTQTGEPKVRSRRAARAAPDTPDPQSLNLEAKQAE-LTAETQKS KAKTKKV
55	orf1ng-1.pep	1140 1150 1160 1170 1180 1190	QADK---DTALAKQREAE TRPATTAFFPRARRARRD-LPQFPQPQPQPQDRLISRYANS
	p45387	1060 1070 1080 1090 1100 1110	RSKRAVFS DPLLQSLFALEAALEVIDAPQQSEKDR LAQEAEKQ-RKQKDLISRYNSA
60	orf1ng-1.pep	1200 1210 1220 1230 1240 1250	LSEFSATLNSVFVAVQDELDRVFAEDRRNAVVTSGIRD TKHYRSQDFRAYRQQ-TDLRQIG
	p45387	1120 1130 1140 1150 1160 1170	LSELSATVNSMLSVQDELDR LFVDQAQSAVWTNIAQDKRRYDSDAFRAYQQQKTNL RQIG
65	orf1ng-1.pep	1260 1270 1280 1290 1300 1310	MQKNLGSGRV GILFSHNRTGNTFDDGIGNSARLAHGAVFCQYIGRFDIGISAGAGFSSG
	p45387	1180 1190 1200 1210 1220 1230	VQKALANGRIGAVFSHSRSDNTFDEQVKNHATLTMMSGFAQYQWGD LQFGVNVGTGISAS

-376-

		1320	1330	1340	1350	1360	1370
5	orf1ng-1.pep	SLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGL					
	p45387	::::   : ::: :::    : :  :  : ::: ::: ::: ::: ::: ::: ::: ::: :::					
		1240	1250	1260	1270	1280	1290
		KMAEEQSRKIHRKAINYGVNASYQFRLGQLGIQPYFGVNRYFIERENYQSEEVVRVKTPSL					
10	orf1ng-1.pep	AFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEW					
	p45387	::: :::   ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
		1300	1310	1320	1330	1340	1350
		AFNRYNAGIRVDYFTFTPDNISVKPYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEV					
15	orf1ng-1.pep	GVNAEIKGFTLSLHAAAAGKQLEAQSAGIKLGYRWX					
	p45387	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::					
		1360	1370	1380	1390		
		GLKAEILHFQISAFISKSQGSQGLGKQNVGVKLGYSRW					

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 20 Example 78

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 655>:

	1	..AAGGTGTGGC	AATTTGTCGA	AGA.CCGCTG	CGTGCCGTCG	TGCCTGCCGA
	51	CAGTTTTGAA	CCGACCGCGC	AAAAATGAA	CCTGTTTAAG	GCGGGTGCGG
25	101	CAACCATTTT	GTTTTATGAA	GATCAAAATG	TCGTCAAAGG	TTTGCAGGAG
	151	CAGTCCCTG	CTTATGCCGC	TAACTTCCCC	GTTGGGCGG	ATCAGGCAAA
	201	CGCGATGGTG	CAGTATGCCG	TTTGGACGAC	ACTTGCCGCG	GTCGGCGTAG
	251	GTGCAAACCT	GCAACATTAC	AATCCCTTGC	CCGATGCGGC	GATTGCCAAA
	301	GCGTGGAATA	TCCCCGAAAA	CTGGTTGTTG	CGCGCACAAA	TGGTTATCGG
	351	CGGTATTGAA	GGGGCGGCAG	GTGAAAAGAC	CTTGAACCC	GTTGCAGAAC
30	401	GTTTGAAAGT	GTTCCGCGCA	TAA		

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

	1	..KVWQFVEXPL	RAVVPADSFE	PTAQKLNLFK	AGAATILFYE	DQNVVKGLQE
	51	QFPAYANFP	VWADQANAMV	QYAVWTTLAA	VGVGANLQHY	NPLPDAAIAK
	101	AWNIPENWLL	RAQMVIGGIE	GAAGEKTFEP	VAERLKVFGA	*

35 Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

	1	..CTGCGTGCCG	TCGTGCCTGC	CGACAGTTTT	GAACCGACCG	CGCAAAAATT
	51	GAACCTGTTT	AAGGCGGGTG	CGGCAACCAT	TTGTTTTAT	GAAGATCAAA
	101	ATGTCGTCAA	AGGTTTGCAG	GAGCAGTTCC	CTGCTTATGC	CGCTAACTTC
40	151	CCCGTTTGGG	CGGATCAGGC	AAACGCGATG	GTGCAGTATG	CCGTTTGGAC
	201	GACACTTGCC	GCGGTGCGCG	TAGGTGCAAA	CCTGCAACAT	TACAATCCCT
	251	TGCCCCGATG	GCGGATTGCC	AAAGCGTGGA	ATATCCCCGA	AACTGGGTTG
	301	TTGCGCGCAC	AAATGGTTAT	CGGCGGTATT	GAAGGGGCGG	CAGGTGAAAA
	351	GACCTTTGAA	CCCGTTGCAG	AACGTTTGAA	AGTGTTCGGC	GCATAA

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

45	1	..LRVVPADSF	EPTAQKLNLF	KAGAATILFY	EDQNVVKGLQ	EQFPAYANFP
	51	PVWADQANAM	VQYAVWTTLA	AVGVGANLQH	YNPLPDAAIA	KAWNIPENWL
	101	LRAQMVIGGI	EGAAGEKTFE	PVAERLKVFG	A*	

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of *N.meningitidis*:

-377-

```

orf6.pep                                KVVWQFVEXPLRAVVPADSF EPTAQKLNLFK
orf6a      QIVEHAVLHTPSSFNSQSARVVVLFGEEDKVVWQFVEDALRAVVPADSF EPTAQKLNLFK
           40      50      60      70      80      90
5
           40      50      60      70      80      90
orf6.pep      AGAATILFYEDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLA AVGVGANLQHY
           |||
orf6a      AGAATILFYEDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLA AVGVGANLQHY
           |||
10
           100     110     120     130     140     150
orf6.pep      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
           |||
15
orf6a      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
           160     170     180     190     200

```

The complete length ORF6a nucleotide sequence <SEQ ID 659> is:

```

1  ATGACCCGTC AATCTCTGCA ACAGGCTGCC GAAAGCCGCC GTTCCATTTA
51  TTCGTTAAAT AAAAATCTGC CCGTCGGCAA AGATGAAATC GTCCAAATCG
20 101  TCGAACACGC CGTTTTGCAC ACACCTTCTT CGTTCAATTC CCAATCTGCC
151  CGTGTGGTCG TGCTGTTTGG CGAAGAGCAT GATAAGGTGT GGCAATTTGT
201  CGAAGACGCG CTGCGTGCCG TCGTGCCTGC CGACAGTTTT GAACCGACCG
251  CGCAAAATTT GAACCTGTTT AAGCGGGGTG CGGCAACTAT TTGTTTTTAT
301  GAAGATCAAA ATGTCGTCAA AGGTTTGAG GAGCAGTTCC CTGCTTATGC
25 351  CGCCAACCTT CCCGTTTGGG CGGACCAGGC GAACGCGATG GTGCAGTATG
401  CCGTTTGGAC GACACTTGCC GCGGTCGGCG TAGGTGCAAA CCTGCAACAT
451  TACAATCCCT TGCCCGATGC GCGGATTGCC AAAGCGTGGA ATATCCCCGA
501  AAATGCGTTG TTGCGCGCAC AAATGTTTAT CGGCGGTATT GAAGGGGCGG
30 551  CAGGTGAAA GACCTTTGAA CCAGTTGCAG AACGTTTGAA AGTGTTCGGC
601  GCATAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 660>:

```

1  MTRQSLQQA ESRRSIYSLN KNLPVGKDEI VQIVEHAVLH TPSSFNSQSA
51  RVVVLFGEEH DKVWQFVEDA LRAVVPADSF EPTAQKLNLF KAGAATILFY
101  EDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLA AVGVGANLQH
35 151  YNPLPDAAIA KAWNIPENWL LRAQMVIIGI EGAAGEKTFE PVAERLKVFG
201  A*

```

ORF6a and ORF6-1 show 100.0% identity in 131 aa overlap:

```

40      50      60      70      80      90      100
orf6a.pep      TPSSFNSQSARVVVLFGEEDKVVWQFVEDALRAVVPADSF EPTAQKLNLFKAGAATILFY
orf6-1      LRAVVPADSF EPTAQKLNLFKAGAATILFY
           10      20      30
45
           110     120     130     140     150     160
orf6a.pep      EDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLA AVGVGANLQHYNPLPDAAIA
orf6-1      EDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLA AVGVGANLQHYNPLPDAAIA
           40      50      60      70      80      90
50
           170     180     190     200
orf6a.pep      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
orf6-1      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
           100     110     120     130
55

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

-378-

	orf6.pep	KVWQFVEXPLRAVVPADSFEPTAQKLNLFK	30
	orf6ng	SNVSLDMSNPTVLRMGLPLYIASLRRGAIYKVWQFVEDALRAVVPADSFEPTAQKLFK	64
5	orf6.pep	AGAATILFYEDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHY	90
	orf6ng	AGAATILFYEDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGAGANLQHY	124
10	orf6.pep	NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGA	140
	orf6ng	NPLPDVAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKVFEPVAERLKVFGA	174

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

15	1	ATGGCCGTTG	CGTCAAATGT	CAGCTTGGAT	ATGTCCAATC	CTACGGTGT
	51	ACGCATGGA	TTACCCCTAT	ATATTGCGTC	CCTAAGAAGG	GGCGCAATAT
	101	ATAAGTGTG	GCAATTTGTC	GAAGACGCGC	TGCGTGCCGT	CGTGCCCTGCC
	151	GACAGTTTG	AACCGACCGC	GCAAAAATTG	AAGCTGTTTA	AGGCGGGCGC
	201	GGCAACCAT	TTGTTTTATG	AAGATCAAAA	TGTCGTCAA	GGTTTGCAGG
	251	AGCAGTCCC	TGCTTATGCC	GCCAACCTTC	CCGTTGGGC	GGACCAGGCG
20	301	AACGCTATG	TACAGTATGC	CGTCTGGACG	ACACTTGCCG	CGGTCGGTGC
	351	AGGTGCAAT	CTGCAACATT	ACAACCCCTT	GCCCGATGTG	GCGATTGCTA
	401	AAGCGTGGAA	TATTCCCGAA	AACTGGCTGT	TGCGCGCGCA	AATGGTTATC
	451	GGTGGTATTG	AAGGGGcggc	aggtgaaaaa	gtctttgaac	CCGTTGCgga
	501	acgtttgAAA	GTGTTTCGGC	CATAA		

25 This encodes a protein having amino acid sequence <SEQ ID 662>:

30	1	MAVASNVSLD	MSNPTVLRMG	LPLYIASLRR	GAIYKVWQFV	EDALRAVVA
	51	DSFEPTAQL	KLFKAGAATI	LFYEDQNVK	GLQEQFPAYA	ANFPVWADQA
	101	NAMVQYAVWT	TLAAGVAGAN	LQHYNPLPDV	AIAKAWNIPE	NWLLRAQMVI
	151	GGIEGAAGEK	VFEPVAERLK	VFGA*		

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

	orf6-1.pep			10	20	30
				LRAVVPADSFEPTAQKLNLFKAGAATILFY		
35	orf6ng	PTVLRMGLPLYIASLRRGAIYKVWQFVEDALRAVVPADSFEPTAQKLFKAGAATILFY				
		20	30	40	50	60
	orf6-1.pep	EDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA				
40	orf6ng	EDQNVVKGLQEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGAGANLQHYNPLPDVAIA				
		80	90	100	110	120
	orf6-1.pep	KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX				
45	orf6ng	KAWNIPENWLLRAQMVIIGGIEGAAGEKVFEPVAERLKVFGAX				
		140	150	160	170	

50 It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

55	1	..GGCTACAACT	ACCTGTTTCGC	GCGCGGCAGC	CGCATCGCCA	ACTACCAAAT
	51	CAACGGCATC	CCCGTTGCCG	ACGCGCTGGC	CGATACGGG	CAATGCCAAC
	101	ACCGCCGCT	ATGAGCGCGT	AGAAGTCGTG	CGCGGCGTGG	CGGGGCTGCT
	151	GGACGGCACG	GGCGAGCCTT	CCGCCACCGT	CAATCTGGTG	CGCAAACGCC
	201	TGACCCGCAA	GCCATTGTTT	GAAGTCCGCG	CCGAAGCGGG	CAACCGcAAA

-379-

251 CATTTCGGGC TGGACGCGGA CGTATCGGGC AGCCTGAACA CCGAAG.crC  
 301 rCTGCGCgGC CGCCTGGTTT CCAcCTTCGG AC GCGGCGAC TCGTGGCGGC  
 351 GCGCGGAACG CAGCCGskAT GCCGAACCT AC GGCATTTT GGAATACGAC  
 401 ATCGCACCGC AAACCGCGT CCACGCArGC ATGGACTACC AGCAGCGGAA  
 451 AGAAACCGCC GACGCGCCGC TCAGcTACGC CGTGACGAC AGCCAAGGTT  
 501 ATGCCACCGC CTTGCGCCCG AAAGACAACC CCGCCACAAA TTGGCGGAAC  
 551 AGCCACCACC GTGCGCTCAA CCTGTTTCGCC GGCATCGAAC ACCGCTTCAA  
 601 CCAAGACTGG AAACCTAAAG CCGAATACGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10 1 ..GYNLYFARGS RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL  
 51 DGTGEPsATV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNTEXX  
 101 LRGRlVSTFG RGDSWRRRER SRXAElyGIL EYDIAPQTRV HAXMDYQQAQ  
 151 ETADAPLSYA VYDSQGYATA FGPKDNPATN WANSHHRLN LFAGIEHRRFN  
 201 QDWKLKAeyD Y..

15 Further work revealed the complete nucleotide sequence <SEQ ID 665>:

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA  
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCCAAACCG CAGGAAAGCA  
 101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC  
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC  
 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC  
 251 GCGACCAAAA CATCAAAACG CTCGACCGCG CCCTGTTGCA GCGGACCGGC  
 301 ACCAGCGGCC AGATTTACGG CTCGACCGCG GCGGGCTACA ACTACCTGTT  
 351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCGCTTG  
 401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC  
 451 GTAGAAGTCG TCGCGGCGCT GCGGGGCTG CTGGACGGCA CGGGCGAGCC  
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCTGACCCGC AAGCCATTGT  
 551 TTGAAGTCG CGCCGAAGCG GGCAACCGCA AACATTTCGG GCTGGACCGC  
 601 GACGTATCGG GCAGCTGAA CACCGAAGGC ACGCTGCGCG GCCGCTGGT  
 651 TTCCACCTTC GGACGCGGCG ACTCGTGCGG GCGGCGCGAA CGCAGCCGCG  
 701 ATGCCGAACT CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC  
 751 GTCCACGCAG GCATGGACTA CCAGCAGCGG AAAGAAACCG CCGACGCGCC  
 801 GTCAGCTAC GCGGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGCGC  
 851 CGAAAGACAA CCCCGCCACA AATTGGGCGA ACAGCGCCA CCGTGCGCTC  
 901 AACCTGTCG CCGGCATCGA ACACCGCTTC AACCAGACT GGAAACTCAA  
 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG  
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC  
 1051 GGTTATTGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTATTGAT  
 1101 CGGCAAATAC CGCTGTTCG GCCGGAACA CGATTAAATC GCGGGTATCA  
 1151 ACGGTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATATCCCC  
 1201 AACCCATT CCAACGCCCTA CGAATTTTCC GCGACGGGTG CCTACCCGA  
 1251 GCCTGCATCG TTGCCCCAAA CCATCCCGCA ATACGGCACC AGGCGGCAAA  
 1301 TCGGCGGCTA TCTCGCCACC CGTTCCGCG CCGCCGACAA CCTTTCGCTG  
 1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACAGCCG  
 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCTACACAG  
 1451 GCATCGTGT CGACCTGACC GGCAACCTGT CTCTTTACGG CTCGTACAGC  
 1501 AGCCTGTTCG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA  
 1551 ACCCGTAACC GGCAACAATC TGGAAGCCGG CATCAAAGGC GAATGGCTTG  
 1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC  
 1651 CTCGCCACCG CAGCAGGAGC CGACCCGAGC GGCAACACCT ACTACCGCGC  
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGGCCGCA  
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC  
 1801 GACCAAGACG GCAGCCGCTT GAACCCGAC AGCGTACCCG AACGCAGCTT  
 1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA  
 1901 CCATCGGCGC AGCGGTGCGC TGGCAGAGCG AAACCCACAC CGACCCTGCC  
 1951 ACGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCG CCGACAACAG  
 2001 CCGCAAAAA GCCTACGCCG TCGCGACAT CATGGCGCGT TACCGCTTCA  
 2051 ATCCGCGCGC CGAAGCTGCG CTGAACGTGG ACAATCTGTT CAACAAACAC  
 2101 TACCGCACCC AGCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA  
 2151 CGCGCGGTTT ACCTATCGGT TTAAATAA

60 This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

1 MTRFKYSLLE AALLPVYAQA DVSVSDPKP QESTELPTIT VTADRTASSN  
 51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG  
 101 TSRQIYSGDR AGYNLYFARG SRIANYQING IPVADALADT GNANTAAYER  
 151 VEVVRGVAGL LDGTGEPsAT VNLVRKRLTR KPLFEVRAEA GNRKHFGLEDA  
 65 201 DVSGSLNTEG TLRGRlVSTF GRGDSWRRER RSRDAELYGI LEYDIAPQTR

-380-

251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL  
 301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HNTAATDLIP  
 351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP  
 401 NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL  
 451 ILGGRYTRYR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS  
 501 SLFVPQSQKD EHGSYLKPVY GNNLEAGIKG EWLEGRNLAS AAVYRARKNN  
 551 LATAAGRDPG GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR  
 601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSEHTDPA  
 651 TLRIIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVDNLFNKH  
 701 YRTQPDHRSY GALRTVNAAF TYRFK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

ORF23 and PupB protein show 32% aa identity in 205aa overlap:

15 Orf23 6 FARGSRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEPSATVNLVRK 65  
 ++RG I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK  
 PupB 215 WSRGFAIQNYEVDGVPTSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK 273  
 20 Orf23 66 RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLVSTFXXXXXXXXXXXXXXXXXAE 125  
 R T + + EAGN +G DVSG L +RGR V+ +  
 PupB 274 RPTAEAQASITGEAGNWDYRGTFDVSGLPTETGNIRGRFVADYKTEKAWIDRYNQSQL 333  
 25 Orf23 126 LYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN 183  
 +YGI E+D++ T + Y + D+PL + S G T N A +W+  
 PupB 334 MYGITFEFDLSEDTLLTVGFSY--LRSDIDSPLRSGLPTRFSTGERTNLKRSNLNAPDWSY 391  
 30 Orf23 184 SHHRALNLFAGIEHRFNQDWKLKAE 208  
 + H + F IE + W K E  
 PupB 392 NDHEQTSFSTSIEQQGLGNGWSGKIE 416

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N. meningitidis*:

35 orf23.pep GYNYLFARGSRIANYQINGIPVADALADTG  
 orf23a QMRDQNIKALDRALLQATGTSRQIYGSDRAGYNYLFARGSRIANYQINGIPVADALADTG  
 90 100 110 120 130 140  
 40 orf23.pep NANTAAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDAD  
 orf23a NANTAAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRETRKPLFEVRAEAGNRKHFGLGAD  
 150 160 170 180 190 200  
 45 orf23.pep VSGSLNTEXXLRGLVSTFGRGDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQAK  
 orf23a VSGSLNAEGTLRGLVSTFGRGDSWRQRERSRDAELYGILEYDIAPQTRVHAGMDYQQAK  
 210 220 230 240 250 260  
 50 orf23.pep ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNLFAGIEHRFNQDWKLKAEYD  
 orf23a ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNLFAGIEHRFNQDWKLKAEYD  
 270 280 290 300 310 320  
 60 orf23.pep Y  
 orf23a YTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTHSASVSLIGKYRLFGRHDLIA  
 330 340 350 360 370 380

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:



-381-

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA  
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCAAAACCG CAGGAAAGCA  
 101 CTGAATTGCC GACCATCACC GTTACCGCGC ACCGCACCGC GAGTTCCAAC  
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGTCGGGC TGCCCATGAC  
 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC  
 251 GCGACCAAAA CATCAAAGCG CTCGACCGCG CCTGTTGCA GCGGACCGGC  
 301 ACCAGCCGCC AGATTACGG CTCCGACCGC GCGGGCTACA ACTACCTGTT  
 351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCGGTTG  
 401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC  
 451 GTAGAAGTCG TGC GCGCGCTG GCGGGGCTG CTGGACGGCA CCGGCGAGCC  
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCCGACCCGC AAGCCATTGT  
 551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTCTG GCTGGGCGCG  
 601 GACGTATCGG GCAGCCTGAA TGCCGAAGGC ACGTGTGCGG GCCGCTGGT  
 651 TTCCACCTTC GGACGCGGCG ACTCGTGGCG GCAGCGCGAA CGCAGCCGCG  
 701 ATGCCGAAT CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC  
 751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC  
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC  
 851 CGAAAGACAA CCCCGCCACA AATTGGGCGA ACAGCCGCCA CCGTGCCTC  
 901 AACCTGTTTC CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA  
 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCTAG  
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC  
 1051 GGTATTGGC ACGCCGACCC GCGCACCAC AGCGCCAGCG TGTCTTAAT  
 1101 CGGCAAATAC CGCCTGTTTC GCGCGAACA CGATTTAATC GCGGGTATCA  
 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC  
 1201 AACGCCATTC CCAACGCCCTA CGAATTTTCC CGCAGGGTG CCTACCCGCA  
 1251 GCCTGTCATCG TTTGCCCAAA CCATCCCGCA ATACGGCACC AGGCGGCAAA  
 1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCCTG  
 1351 ATACTCGGCG GCAGATACAG CCGTTACCGC ACCGGCAGCT ACGACAGCCG  
 1401 CACACAAGGC ATGACCTATG TGTCGGCCAA CCGTTTCACC CCCTACACAG  
 1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CGCTTTACGG CTCGTACAGC  
 1501 AGCCTGTTTC TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA  
 1551 ACCCGTAACC GGCAACAATC TGAAGCCGG CATCAAAGGC GAATGGCTTG  
 1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC  
 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC  
 1701 QGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGCGCCGCA  
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC  
 1801 GACCAAGACG GCAGCCGCTT GAACCCGAC AGCGTACCCG AACGCAGCTT  
 1851 CAAACTCTTC ACTGCTTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA  
 1901 CCATCGGCGC AGGCGTGGCG TGGCAGAGCG AAACCCACAC CGACCTGCC  
 1951 ACGTCCGCA TCCCCAACCC CGCCGCCAAA GCGCGCGCCG CCGACAACAG  
 2001 CCGCCAAAAA GCCTACGCGG TCGCCGACAT CATGGCGCGT TACCGCTTCA  
 2051 ATCCGCGCGC CGAAGTGTG CTGAACGTGG ACAATCTGTT CAACAAACAC  
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA  
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

45 This encodes a protein having amino acid sequence <SEQ ID 668>:

1 MTRFKYSLLF AALLPVYAQA DVSVDDBPKP QESTELPTIT VTADRTASSN  
 51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKA LDRALLQATG  
 101 TSRQIYGS DR AGYNLFFARG SRIANYQING IPVADALADT GNANTAYER  
 151 VEVVRGVAGL LDGTGEP SAT VNLVRKRPR KPLFEVRAEA GNRKHFGLGA  
 201 DVSGSLNAEG TLRGRVSTF GRGDSWRQRE RSRDAELYGI LEYDIAPQTR  
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL  
 301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HNTAATDLIP  
 351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP  
 401 NAIPNAYEFS RTGAYPQFAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL  
 451 ILGGRYSRYP TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS  
 501 SLFVPQSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRNLNAS AAVYRARKNN  
 551 LATAAGRDPS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR  
 601 DQDGSRLNPD SVPERSEKLF TAYHEAPEAP SGWTIGAGVR WQSETHDPA  
 651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNDLNFKNH  
 701 YRTQPRHSY GALRTVNAAF TYRFK\*

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

		10	20	30	40	50	60
orf23a.pep		MTRFKYSLLF	AALLPVYAQA	DVSVDDBPKP	QESTELPTIT	VTADRTASSN	DGYTVSGTHT
65 orf23-1		MTRFKYSLLF	AALLPVYAQA	DVSVDDBPKP	QESTELPTIT	VTADRTASSN	DGYTVSGTHT
		10	20	30	40	50	60

-382-

		70	80	90	100	110	120
	orf23a.pep	PLGLEPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNYLFARG					
5	orf23-1	PLGLEPMTLREIPQSVSVITSQQMRDQNIKTLDRALLQATGTSRQIYGSDRAGYNYLFARG					
		70	80	90	100	110	120
	orf23a.pep	130	140	150	160	170	180
10	orf23-1	SRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRPTR					
		130	140	150	160	170	180
	orf23a.pep	190	200	210	220	230	240
15	orf23-1	KPLFEVRAEAGNRKHFGGLGADVSGSLNAEGTLRGRLVSTFGRGDSWRQRERSRDAELYGI					
		190	200	210	220	230	240
	orf23a.pep	250	260	270	280	290	300
20	orf23-1	LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL					
		250	260	270	280	290	300
	orf23a.pep	310	320	330	340	350	360
25	orf23-1	NLFAGIEHFRNQDWKLKAEYDYTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH					
		310	320	330	340	350	360
	orf23a.pep	370	380	390	400	410	420
30	orf23-1	SASVSLIGKYRLFGREHDLIAGINGKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
		370	380	390	400	410	420
	orf23a.pep	430	440	450	460	470	480
35	orf23-1	FAQTIPQYGTTRRQIGGYLATRFRAADNLSLILGGRYSTRYRTGSYDSRTQGMTYVSANRFT					
		430	440	450	460	470	480
	orf23a.pep	490	500	510	520	530	540
40	orf23-1	PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVGTGNNLEAGIKGEWLEGRNAS					
		490	500	510	520	530	540
	orf23a.pep	550	560	570	580	590	600
45	orf23-1	AAVYRARKNNLATAAGRDPGSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR					
		550	560	570	580	590	600
	orf23a.pep	610	620	630	640	650	660
50	orf23-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPATLRIPNPAAK					
		610	620	630	640	650	660
	orf23a.pep	670	680	690	700	710	720
55	orf23-1	ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDRHSYGALRTVNAAF					
		670	680	690	700	710	720
	orf23a.pep	TYRFXK					
60	orf23-1	TYRFXK					
		TYRFXK					
65							
70							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N. gonorrhoeae*:

5	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLD	51
	orf23ng	SAVDACRIPGYNYLFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPD	60
10	orf23.pep	GTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLRVSTFGR	111
	orf23ng	GTGEPSATVNLVRKHPTKPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGLRVSTFGR	120
15	orf23.pep	GDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAF	180
	orf23.pep	GPKDNPATNWSNSRNRALNLFAGIEHFRNQDWKLKAEYDY	211
	orf23ng	GPKDNPATNWSNSRNRALNLFAGIEHFRNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising  
20 amino acid sequence <SEQ ID 670>:

	1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
	51	VVRGVAGLPD	GTGEPSATVN	LVRKHPTKRP	LFEVRAEAGN	RKHFGLGADV
	101	SGSLNAEGTL	RGLRVSTFGR	GDSWRQLERS	RDAELYGILE	YDIAPQTRVH
25	151	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNSRNRALNL
	201	FAGIEHFRNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
	251	WHADPRTHSA	SMSLTGKYRL	FGREHDLIAG	INGYKYASNK	YGERSIIPNA
	301	IPNAYEFSRT	GAYPQPSSFA	QTIPQYDTRR	QIGGYLATRF	RAADNLSLIL
	351	GGRYSRYRAG	SYNSRTQGMT	YVSANRFTPY	TGIVFDLTGN	LSLYGSYSSL
30	401	FVPQLQKDEH	GSYLKPVGTG	NLEADIKGEW	LEGRLNASAA	VYRARKNNLA
	451	TAAGRDQSGN	TYYRAANQAK	THGWEIEVGG	RITPEWQIQ	QYGSQSKPRDQ
	501	DGSRNLPDSV	PERSFKLFTA	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
	551	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTLSLN	VDNLFNKHRY
	601	TQPDHRHSGA	LRTVNAAFTY	RFK*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
	101	CCGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGCACCGC	GAGTTCCAAC
	151	GACGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCCGGC	TGCCCATGAC
40	201	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
	251	CGGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTTGCA	GGCGACCGGC
	301	ACCAGCCGCC	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAACCTACCA	AATCAACGGC	ATCCCGTGTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
45	451	GTAGAAGTCG	TGCGCGGCGT	GGCGGGGCTG	CCGGACGGCA	CGGGCGAGCC
	501	TTCTGCCACC	GTCAATCTGG	TACGCAAACA	CCCGACCCGC	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTCCG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCTGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGGCG	GCAGCTCGAA	CGCAGCCGCG
	701	ATGCCGAACT	CTACGGCATT	TTGGAATACG	ACATCGCACC	GCAAACCCGC
50	751	GTCCACGCAG	GCATGACTA	CCAGCAGCGC	AAAGAAACCG	CAGACGCGCC
	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
	851	CAAAAGACAA	CCCCGCCACA	AATTGGTCTG	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTCG	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGTTGTGG
55	1001	CAGGCGTACT	TTCCATCGAC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCGATCC	GCGCACCAC	AGCGCCAGCA	TGTCATTGAC
	1101	CGGCAAATAC	CgcctGTTCC	GCCGCGAGCA	CGATTTAATC	GCGGGTATCA
	1151	ACGGCTACAA	ATACGCCAGC	AACAAATACG	GCGAACGCAG	CATCATTCCC
	1201	AACGCCATT	CCAACGCCTA	CGAATTTTCC	CGCACGGGCG	CCTATCCGCA
60	1251	GCCATCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGACACC	AGGCGGCAAA
	1301	TCGGCGGCTA	TCTCGCCACC	CGTTTCCGCG	CCGCCGACAA	CCTTTCGCTG
	1351	ATACTCGGCG	GCAGATACAG	CCGCTACCGC	GCAGGCAGCT	ACAACAGCCG

-384-

1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCCTACACAG  
 1451 GCATCGTGTT CGATCTGACC GGCAACCTGT CGCTTTACGG CTGCTACACG  
 1501 AGCCTGTTTC TCCCGCAATT GCAAAAAGAC GAACACGGCA GCTACCTGAA  
 1551 ACCCGTAACC GGCAACAATC TGGGAAGCCGA CATCAAAGGC GAATGGCTTG  
 1601 AAGGGCGTCT GAACGCATCC GCCGCGGTGT ACCGCGCCCG TAAAAACAAC  
 1651 CTCGCCACCG CAGCAGGACG CGACCAGAGC GGCAACACCT ACTATCGCGC  
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA  
 1751 TCACGCGCGA ATGGCAGATA CAGGCAGGCT ACAGCCAAAG CAAACCCCGC  
 1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTAcCCG AACGCAGCTT  
 1851 CAAACTCTTC ACCGCCTACC ACTTAGCCCC CGAAGCCCCC AGCGGCCCGA  
 1901 CCATcggTGC GGGTGTGCGC CGGCAGGGCG AAACCCACAC CGACCCAGCC  
 1951 GCGCTCCGCA TCCCAACCC CGCCGCCAAA GCCCGCGCCG TCGCCAACAG  
 2001 CCGCCAGAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA  
 2051 ATCCGCGCAC CGAAGTCTCG CTGAACGTGG ACAACCTGTT CAACAACAC  
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA  
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

1 MTRFKYSLLF AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN  
 51 DGYTVSGTHT PFGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG  
 101 TSRQIYGSDR AGYNLYFARG SRIANYQING IPVADALADT GNANTAAYER  
 151 VEVVRGVAGL PDGTGEP SAT VNLVRKHPT R KPLFEVRAEA GNRKHFGLGA  
 201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQLE RSRDAELYGI LEYDIAPQTR  
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWSNSRNRAL  
 301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HSTAATDLIP  
 351 GYWHADPRTH SASMSLTGKY RLFGREHDLI AGINGYKYAS NKYGERSTIP  
 401 NAIPNAYEFS RTGAYPQPSS FAQTIPQYDT RRQIGGYLAT RFRAADNLSL  
 451 ILGGYSRYR AGSYNSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS  
 501 SLFVPQLQKD EHGSYLKPV T GNNLEADIKG EWLEGRNAS AAVYRARKNN  
 551 LATAAGRDQS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKER  
 601 DQDGSRLNPD SVPERSFKLF TAYHLAPEAP SGRTIGAGVR RQGETHTDPA  
 651 ALRIPNPAK ARAVANSRQK AYAVADIMAR YRFNPRTLS LNVNDLNFKNH  
 701 YRTQPDHRSY GALRTVNAAF TYRKF\*

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

		10	20	30	40	50	60
35	orf23-1.pep	MTRFKYSLLEAALLPVYAQADSVSDDPKPQESTELPTITVTADRTASSNDGYTVSGTHT					
	orf23ng-1	MTRFKYSLLEAALLPVYAQADSVSDDPKPQESTELPTITVTADRTASSNDGYTVSGTHT					
		10	20	30	40	50	60
40	orf23-1.pep	PLGLPMTLREIPQSVSVITSQQMRDQNIKTLDRLALQATGTSRQIYGS DRAGYNLYFARG					
	orf23ng-1	PFGLPMTLREIPQSVSVITSQQMRDQNIKTLDRLALQATGTSRQIYGS DRAGYNLYFARG					
		70	80	90	100	110	120
45	orf23-1.pep	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTR					
	orf23ng-1	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPDGTGEP SATVNLVRKHPT R					
50		130	140	150	160	170	180
	orf23-1.pep	KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGRLVSTFGRGDSWRRRERSRDAELYGI					
	orf23ng-1	KPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGRLVSTFGRGDSWRQLERSRDAELYGI					
55		190	200	210	220	230	240
	orf23-1.pep	LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWSNSRNRAL					
	orf23ng-1	LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWSNSRNRAL					
60		250	260	270	280	290	300
	orf23-1.pep	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH					
	orf23ng-1	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH					
65		310	320	330	340	350	360

-385-

		310	320	330	340	350	360
		370	380	390	400	410	420
5	orf23-1.pep	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
	orf23ng-1	SASMSLTGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPSS					
		370	380	390	400	410	420
		430	440	450	460	470	480
10	orf23-1.pep	FAQTIPQYGTTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT					
	orf23ng-1	FAQTIPQYDTRRQIGGYLATRFRAADNLSLILGGRYSTRYRAGSYNSRTQGMTYVSANRFT					
		430	440	450	460	470	480
		490	500	510	520	530	540
15	orf23-1.pep	PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVGTGNNLEAGIKGEWLEGRNAS					
	orf23ng-1	PYTGIVFDLTGNLSLYGSYSSLFVQLQKDEHGSYLKPVGTGNNLEADIKGEWLEGRNAS					
		490	500	510	520	530	540
		550	560	570	580	590	600
	orf23-1.pep	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRTPEWQIQAGYSQSKTR					
25	orf23ng-1	AAVYRARKNNLATAAGRDQSGNTYYRAANQAKTHGWEIEVGGRTPEWQIQAGYSQSKPR					
		550	560	570	580	590	600
		610	620	630	640	650	660
	orf23-1.pep	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPAALRIPNPAK					
30	orf23ng-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGRTIGAGVRWQSEHTDPAALRIPNPAK					
		610	620	630	640	650	660
		670	680	690	700	710	720
	orf23-1.pep	ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRHSYGALRTVNAAF					
35	orf23ng-1	ARAVANSRQKAYAVADIMARYRFNPRTLSLNVDNLFNKHYRTQPDHRHSYGALRTVNAAF					
		670	680	690	700	710	720
40	orf23-1.pep	TYRFXK					
	orf23ng-1	TYRFXK					

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

45	sp P16869 FHUE_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR >gi 1651542 gnl PID d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli]
	>gi 1651545 gnl PID d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1787344 (AE000210) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrulic acid precursor [Escherichia coli] Length = 729
50	Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
55	Query: 38 TITVTADRTASSN--DGYTVSGTHTPFGLPMTLREIPQSVSVITSQQMRDQNIKTLDRL 95 T+ V TA + + Y+V+ T + MT R+IPQSV++++ Q+M DQ ++TL + Sbjct: 43 TVIVEGSATAPDDGENDYSVTSTSAGTKMQMTQRDIPQSVTIVSQQRMEDQQLQTLGEVM 102
60	Query: 96 LQATGTSRQIYGSDRAGYNLYFARGSRANYQINGIP-----VADALADTGNANTAA 147 G S+ SDRA Y ++RG +I NY ++GIP + DAL+D A Sbjct: 103 ENTLGISKSQADSDRALY---YSRGFQIDNYMVDGIPTYFESRWNLGDALSDM-----AL 154
65	Query: 148 YERVEVVRGVAGLPDGTGEP SATVNLVRKHPTKPLF-EVRAEAGNRKHFGGLGADVSGSL 206 +ERVEVVRG GL GTG PSA +N+VRKH T + +V AE G+ AD+ L Sbjct: 155 FERVEVVRGATGLMTGTGNPSAAINMVRKHATSREFKGDVSAEYGSWNKERYVADLQSP 214
	Query: 207 NAEGLRGRVLSTFGRGDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQOAKETADA 266 +G +R R+V + DSW S GI++ D+ T + AG +YQ+ + Sbjct: 215 TEDGKIRARIVGGYQNNDSWLDRYNSEKTFESGIVDADLGDLTLSAGYEYQRIDVNSPT 274
70	Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWSNSRNRALNLEAGIEHRFNQDWKLKAEYDYTRSR 326

-386-

+++ G + ++ + A +W+ + +F ++ +F W+ ++  
 Sbjct: 275 WGGLPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVEMTLKQQFADTWQATLNATHSEVE 334  
 Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHSA-SMSLTGKYRLF 374  
 F + Y A V D ++ PG+ W++ R A + G Y LFG  
 Sbjct: 335 FDSKMMYVDAYVNKADGMLVGPYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSYELFG 394  
 Query: 375 REHDLIAGINGYKYASNKYGER--SIIPNAIPNAYEFSRTGAYFPQSSFAQTIPQYDTRR 432  
 R+H+L+ G Y +N+Y +I P+ I + Y F+ G +EQ Q++ Q DT  
 Sbjct: 395 RQHNLMFQ-GSYSKQNNRYFSSWANIFPDEIGSFYNFN--GNFPQTDWSPQSLAQDDTTH 451  
 Query: 433 QIGGYLATRFRAADNLSLILGGYSRYRAGSYNSRTQGMY-VSANRFTPYTGIVFDXXX 491  
 Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD  
 Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMKHNHTPYAGLVFDIND 504  
 Query: 492 XXXXXXXXXXXXFPQLQKDEHGSYLKPVGTNNLEADIKGEWLEGRNLSAAVYRARKNNL 551  
 F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+  
 Sbjct: 505 NWSTYASYTSIFPQNDRDSSGKYLAPITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564  
 Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRTPEWQIQAGYSQSKPRDQDGSRLN 608  
 A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N  
 Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGTVSKGVFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624  
 Query: 609 PDSVPERSEKLEFAYHLAEPASGRTIGAGVRRQGETHTDPAALRIPNPAKARAVANSR 668  
 P ++P + K+ET+Y L P P T+G GV Q +TD P RA  
 Sbjct: 625 P-NLPRTTVKMFTSYRL-PVMPE-LTVGGGVNWQNRVYTDIV-----TPYGTFR-----E 672  
 Query: 669 QKAYAVADIMARYRFNPRTELSLNVNLFNKHYRTQPDHRH-SYGALRTVNAAFTYRF 724  
 Q +YA+ D+ RY+ L NV+NLF+K Y T + YG R + TY+F  
 Sbjct: 673 QGSYALVDLFTRYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGTPRNFSITGTYYQF 729

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC  
 51 GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA  
 101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC  
 151 AGCGTCAGcA CGCCTGCTTC GGCGgcGgCa ATCATACCTT CGTCTTCGGA  
 201 AACGGGGATA AACGcGCCAC TCAAACCCCC GACCGCGCTG GAAGCCATCA  
 251 TGCCGCGCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG  
 301 CCGTGCCTAC CGCAGACGCT CAAGCCCATT ThTTCAAGAA TGCGTGCCAC  
 351 TnAGTCGCGG ACGGGG..

This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS  
 51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPFFFTA SFSNAKAAV  
 101 PCVPQTLKPI XSRMRATXSP TG..

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

```

1   ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
151 AGCGTCAGCA CGCCTGCTTC GCGGCGGGCA ATCATACCTT CGTCTTCGGA
201 AACGGGGATA AACGCGCCAC TCAAACCCCG GACCGCGCTG GAAGCCATCA
251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301 CCGTGCGTAC CGCAGACGCT CAAGCCCATT TCTTCAAGAA TGCCTGCCAC
351 TGAGTCGCGG ACGGCGGGGG TCGGCGCCAG CGACAAGTCG AGAATACCAA
401 ACGGGATATT CAGCATTTT T GAGGCTTCGC GGCCGATGAG TTCGCCCACG
451 CGGGTAATTT TGAAAGCAGT TTTCTTCACT ACTTCCGCAA CTTCGGTCAA
501 TGTCGTTGCA TCTGAATTT CCAACGCGGC TTTTACGACA CCTGGGCCGG
551 ATACGCGGAC ATTGATAACG GCATCCGCTT CGCCCGAACC ATGAAACCGG
601 CCCGCCATAA ACGGGTTGTC TTCCACCGCG TTGCAGAAAC CGACAATTTT
15  651 AGCGCAGCCG AAACCTTCGG GCGTGATTTT CGCCGTGCGT TTGACGGTTT
701 CGCCCGCCAG CTTGACCGCA TCCATATTGA TACCGGCACG CGTACTGCCG
751 ATATTGATGG AGCTGCACAC AATATCGGTA GTCTTCATCG CTTCGGGAAT
801 GGAGCGGATT AACACCTCAT CCGAAGGCGA CATCCCTTTT TGCACCAACG
851 CGGAAAACCC GCCGATAAAA GACACACCGA TGGCTTTGGC AGCTTTATCC
20  901 AAAGTTTGCG CCACGCTGAC GTAA

```

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

```

1   MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS
51  SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
25  101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSTP
151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDTPTLIT ASASPEP*NA
201 PAINGLSSTA LQNTTILAQP KPSGVISAVR LTVSPASLTA SILIPARVLE
251 ILMELHTISV VFIASGMERI NTSSEGDIPF CTNAEKPPIK DTPMALAALS
301 KVCATLT*

```

Computer analysis of this amino acid sequence gave the following results:

### 30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N. meningitidis*:

```

35  orf24a.pep      10      20      30      40      50      60
      MRTAVVLLLIMFMAASSAMMPMV CAGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf24  MRTAVVLLLIMFMAASSAMMPMV CAGVSPGTAIISXPTEQTAVMASSLSNVSTPASAAA
      10      20      30      40      50      60

40  orf24a.pep      70      80      90      100     110     120
      IIPSSSXTGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf24  IIPSSSXTGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP
      70      80      90      100     110     120

45  orf24a.pep     130     140     150     160     170     180
      TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf24  TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
      130     140     150     160     170     180

50  orf24a.pep     190     200     210     220     230     240
      PGPDTPTLITASASPEPKNAPAI XGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf24  PGPDTPTLITASASPEPKNAPAI NGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA
      190     200     210     220     230     240

55  orf24a.pep     250     260     270     280     290     300
      SILIPARVLPILMELHTISVVFIASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf24  SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS
      250     260     270     280     290     300

```

-388-

```

orf24a.pep  KVCATLTX
            |||||
orf24       KVCATLTX

```

5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

```

1  ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG TGTGTCGCCG GGAACGGCAA
101 TCATATCCAA NCCGACCGAA CAAACGGCGG TCATCGCTTC GAGTTTATCC
151 AACGTCAGCA CGCCTGCTTC GGCGGCGGCA ATCATACCTT CGTCTTCGGA
10  201 NACGGGGATA AACCGGCCAC TCAAACCGCC AACCGCGCTC GAAGCCATCA
251 TGCGGCCCTT TTTCAGCGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301 CCGTGCGTAC CGCAGACGCT CAAACCCATT TCTTCAAGAA TGCGCGCCAC
351 CGAGTCGCCG ACGGCAGGGG TCGGTGCCAG CGACAAGTCG AGAATACCAA
15  401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TTCGCCACG
451 CGGGTAATTT TGAAGCGGTT TTTCTTCACA ACTTCGGCAA CTTCGGTCAA
501 TGTCGTTGCA TCCGAATTTT CCAACGCGGC TTTTACGACA CCCGGGCCGG
551 ATACGCCGAC ATTAATCACA GCATCCGCTT CGCCTGAGCC GTGAAACGCG
601 CCCGCCATAN ACGGGTTGTC TTCCNCCGCG TTGCAGACA CGACGATTTT
651 GGCGCAGCCG AAACCTTCTA GTGTGATTTC ANCCGTGCGT TTGATGGTTT
20  701 CGCCCGCCAG TCTGACCGCG TCCATATTGA TACCGGCGCG CGTACTGCCG
751 ATATTGATGG AGCTGCACAC GATATCAGTA GTCTTCATCG CTTCGGGAAT
801 GGAACGGATN AACACCTCGT CAGAAGCGCA CATACTTTT TGCACCAGCG
851 CGGAAAAGCC GCCAATAAAA GACACGCCGA TGGCTTTGGC AGCCTTATCC
901 AAAGTTTGCG CCACGCTGAC GTAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 678>:

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS
51  NVSTPASAAA IIPSSSXTGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAV
101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSPT
151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDTPTLIT ASASPEP*NA
201 PAIXGLSSXA LQNTTILAQP KPSSVISXVR LMVSPASLTA SILIPARVLP
251 ILMELHTISV VFISGEMERX NTSSEGDIPF CTSAEKPPIK DTPMALAALS
301 KVCATLT*

```

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

```

35  orf24a.pep      10      20      30      40      50      60
      MRTAVVLLLIMPMMAASSAMMPMVCAVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA
      orf24-1      10      20      30      40      50      60
      MRTAVVLLLIMPMMAASSAMMPMVCAVSPGTAIISKTEQTAVMASSLSVSTPASAAA
40  orf24a.pep      70      80      90      100     110     120
      IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAAVPCVPQTLKPISSRMRATESP
      orf24-1      70      80      90      100     110     120
      IIPSSSETGINAPLKPPTALEAIMPPFFTASFSNAKAAVPCVPQTLKPISSRMRATESP
45  orf24a.pep      130     140     150     160     170     180
      TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
      orf24-1      130     140     150     160     170     180
      TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
50  orf24a.pep      190     200     210     220     230     240
      PGPDTPTLITASASPEPXNAPAIIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
      orf24-1      190     200     210     220     230     240
      PGPDTPTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA
55  orf24a.pep      250     260     270     280     290     300
      SILIPARVLPILMELHTISVVFISGEMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS
      orf24-1      250     260     270     280     290     300
      SILIPARVLPILMELHTISVVFISGEMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS

```



orf24a.pep KVCATLTX  
 5 orf24-1 KVCATLTX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from *N.gonorrhoeae*:

10	orf24.pep	MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIISKPTAQAVMASSLSVSTPASAAA	60
	orf24ng	MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIMSKPTAQAVMASSLSVSTPASAAA	60
15	orf24.pep	IIPSSSETGINAPLKPPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPISSRMRTXSP	120
	orf24ng	IIPSSSETGINAPLKPPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPISSRMRTATESP	120
	orf24.pep	TG	122
20	orf24ng	TAGVGASDKSRMPNGIFSIFEASRPMSPTRVILKAVFFTTTSATSVRLTASEFSSAALT	180

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

	1	ATGCGCACGG	CGGTGGTTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
	51	GGCGATGATG	CCGGAAATGG	TGTGCGCGGG	CGTGTGCGCG	GGAACGGCAA
25	101	TCATGTCCAA	ACCAACGGAG	CAGACGGCGG	TCATGGCTTC	GAGTTTGTCC
	151	AGCGTCAACA	CGCCTGCCTC	GGCGGCGGCA	ATCATACCTT	CGTCTTCGGA
	201	AACGGGGATA	AACGCGCCGC	TCAAACCGCC	GACCGCGCTG	GAAGCCATCA
	251	TGCGGCCCTT	TTTCACGGCA	TCGTTTACGA	ATGCCAAAGC	TGCTGTTGTG
	301	CCGTGCGTAC	CGCAGACGCT	CAAGCCCAT	TCTTCAAGAA	TGCGGCGCCAC
	351	CGAGTCGCGG	ACGGCGGGGG	TCGGTGCCAG	CGACAAATCG	AGAATGCCGA
30	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	GACCGATGAG	TTCGCCACG
	451	CGGTGATTT	TGAAAGCGGT	TTTCTTCACG	ACTTCGGCGA	CCTCGGTACG
	501	GCTGACCGCG	TCCGAATTTT	CCAGCGCGGC	TTTGACCACG	CCTGGACCGG
	551	ATACGCCGAC	ATTAATCACA	GCATCCGCTT	CGCCCGAGCC	GTGGAACGCA
	601	CCCGCCATAA	ACGGATTGTC	TTCCACCGCG	TTGCAGAACA	CGACGATTTT
35	651	GGCGCAGCCG	AAACCTTCGG	GTGTGATTTC	AGCCGTGCGT	TTGATGGTTT
	701	CGCTGCCAG	CTTGACCGCA	TCCATATTGA	TACCGGCACG	CGTGCTGCCG
	751	ATATTGATGG	AGCTGCACAC	GATATCGGTA	GTTTTTCATG	CTTCGGGAAC
	801	GGAACGGATC	AACACCTCAT	CCGAAGGCGA	CATACCTTTT	TGCACCGACG
40	851	CGGAAAAGCC	GCCGATAAAG	GACACGCCGA	TGGCTTTGGC	TGCCTTGTC
	901	AAAGTCTGCG	CCACGCTGAC	ATAA		

This encodes a protein having amino acid sequence <SEQ ID 680>:

	1	MRTAVVLLLI	MPMAASSAM	PEMVCAGVSP	GTAIMSKPTE	QTAVMASSLS
	51	SVNTPASAAA	IIPSSSETGI	NAPLKPPPTAL	EAIMPPFFTA	SFSNAKAAV
45	101	PCVPQTLKPI	SSRMRTATESP	TAGVGASDKS	RMENGIFSIF	EASREMSPT
	151	RVILKAVFFT	TSATSVRLTA	SEFSSAALT	PGPDTPTLIT	ASASPEPWNA
	201	PAINGLSSTA	LQNTTILAQP	KPSGVISAVR	LMVSPASLTA	SILIPARVLP
	251	ILMELHTISV	VFIASGTERI	NTSSEGDIPF	CTSAEKPIK	DTPMALAALS
	301	KVCATLT*				

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

50	orf24-1.pep	10	20	30	40	50	60
	orf24ng	10	20	30	40	50	60
55	orf24-1.pep	70	80	90	100	110	120
	orf24ng	70	80	90	100	110	120

-390-

		130	140	150	160	170	180
	orf24-1.pep	TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAFTT					
5	orf24ng	TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVRLTASEFSSAALT					
		130	140	150	160	170	180
	orf24-1.pep	PGPDTPTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
10	orf24ng	PGPDTPTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTMVSPASLTA					
		190	200	210	220	230	240
	orf24-1.pep	SILIPARVLPILMELHTISVVFIA SGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
15	orf24ng	SILIPARVLPILMELHTISVVFIA SGTERINTSSEGDIPFCTSAEKPPIKDTPMALAALS					
		250	260	270	280	290	300
20	orf24-1.pep	KVCATLTX					
	orf24ng	KVCATLTX					

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

30	1	..ACCGACGTGC	AAAAAGAGTT	GGTCGGCGAA	CAACGCAAGT	GGGCGCAGGA
	51	AAAAATCAGC	AACTGCCGAC	AAGCCGCCGC	GCAGGCAGAC	CGGCAGGAAT
	101	ACGCCGAATA	CCTCAAGCTG	CAATGCGACA	CGCGGATGAC	GCGCGAACGG
	151	ATACAGTATC	TTCGCGGCTA	TTCCATCGAT	TAG	

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

35	1	..TDVQKELVGE	QRKWAQEKIS	NCRQAAQAD	RQEYAEYLKL	QCDTRMTRE
	51	IQYLRGYSID	*			

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

	1	ATGTATCGGA	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCCGCTTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCATTGGA	ATGCGCCAAC	CCCGCCGTGT
40	101	TGCAAGGCAT	ACGCGGCAAT	ATTCAGGAAA	CGCTCACGCA	GGAAGCGCGT
	151	TCTTTCGCGC	GCGAAGACGG	CAGGCAGTTT	GTGATGCCG	ACAAAATTAT
	201	CGCCGCCGCC	TACGGTTTGG	CGTTTTCTTT	GGAACACGCT	TCGGAACGCG
	251	AGGAAGGCGG	GCGCACGTTT	TGTATCGCCG	ATTTGAACAT	TACCGTGCCG
	301	TCTGAAACGC	TTGCCGATGC	CAAGGCAAAC	AGCCCCCTGT	TGTACGGGGA
45	351	AACTGCTTTG	TCGGATATTG	TGCGGCAGAA	GACGGGCGGC	AATGTCGAGT
	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCCTGCC	CGTCAAGAC
	451	GGTCAGACGG	CATTTGTCGA	CAACACGGTC	GGTATGGCGG	CGCAAACGCT
	501	GTCTGCCGCG	CTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
	551	GCAAGGCGGT	GAAAAAGAA	GACGCGGTCA	GGATTTTGAG	CGGAAAGGCC
50	601	CGTGAAGAAG	AACCGTCCAA	ACCCACGCC	GAAGACATT	TGGAACACAA
	651	TGCCGCCGCG	GGCGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGGCCCGG
	701	AACCGGAAAT	CCTGCATCCT	GACGACGGCG	AGCGTGCCGA	TACCGTTACC
	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGGCG	GTACAAAACC	AGCGTGCGGA
	801	ATCCGAAATT	ACCAAACCTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
55	851	AGTTGGTCGG	CGAACACGCG	AAGTGGGCGC	AGGAAAAAAT	CAGCAACTGC
	901	CGACAAGCCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
	951	GCTGCAATGC	GACACGCGGA	TGACGCGCGA	ACGGATACAG	TATCTTCGCG
	1001	GCTATTCCAT	CGATTAG			

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```

      1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQGIRGN IQETLTQEAR
     51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
    101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
    151 GQTAFVDNTV GMAAQTLCAA LLPYGVKSIV MIDGKAVKKE DAVRILSGKA
    201 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEFEILHP DDGERADTVT
    251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC
    301 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

```

Computer analysis of this amino acid sequence gave the following results:

#### 10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N. meningitidis*:

```

      10      20      30
    orf25.pep      TDVQKELVGEQRKWAQEKISNCRQAAAQAD
    15      |||||
    orf25a      VTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEXRKWAQEKISNCRQAAAQAD
      250      260      270      280      290      300

      40      50      60
    orf25.pep      RQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
    20      |||||
    orf25a      RQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
      310      320      330

```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

```

    25      1 ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
      51 CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
    101 TGCAANGCAT ACGCNGCAAT ATTCAGGAAA CGCTCAGCA GGAAGCGCGT
    151 TCTTTCGCGC GCGAAGACNG CANGCAGTTT GTCGATGCCG ACNAAATTAT
    201 CGCCGCGGCC TANGNTNNGN NGNTNTCTTT GGAACACGCT TCGGAAACGC
    301 AGGAAGGCGG GCGCACGTTC TGTNTCGCCG ATTTGAACAT TACCGTGC CG
    351 TCTGAAACGC TTGCCGATGC CAAGGCAAAC AGCCCCCTGC TGTACGGGGA
    401 AACCCTTTTG TCGGATATTG TCGGCGAGAA GACGGGCGGC AATGTCGAGT
    451 TTAAGAGACG CGTATTGACG GCAGCCGTCC GCTTCTTACC CGTCAAAGAC
    501 GGTCAGANGG CATTTGTGCA CAACACGGTC GGTATGGCGG CGCAAACGCT
    551 GTCTGCCGCG TTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
    601 GCAAGGCGGT AAAAAAGAA GACGCGGTCA GGATTNTGAG CNGANAAGCC
    651 CGTGAANAAG AACCCTCCAA ANCCNNGCCC GAAGACATTT TGGAAACATA
    701 TGCCGCCGGA GGGGATGCAG ACGTACCCCA AGCCGAGAGG GACGCGCCCG
    751 AACCAGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
    801 GTATCACGGG GCGAAGTGA AGAGGCGCGN GTACAAAACC AGCGTGCGGA
    851 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
    901 AGTTGGTCCG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
    951 CGACAAGCCG CCGCGCAGGC AGACGGGCGG GAATACGCCG AATACCTCAA
    1001 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
      GCTATTCCAT CGATTAG

```

This encodes a protein having amino acid sequence <SEQ ID 686>:

```

      1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
     51 SFAREDXXQF VDADXIIAAA XXXXSLEHA SETQEGGRTF CXADLNITVP
    101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
    151 GQXAFVDNTV GMAAQTLCAA LLPYGVKSIV MIDGKAVKKE DAVRIXSXXA
    201 REXEPSKXXP EDILEHNAAG GDADVQAGE DAPEFEILHP DDGERADTVT
    251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEXR KWAQEKISNC
    301 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

```

    55      10      20      30      40      50      60
    orf25a.pep MYRKLIALPFALLLAACGRE EPPKALECANPAVLQXIRXNIQETLTQEARSFAREDXXQF
    orf25-1      |||||
    MYRKLIALPFALLLAACGRE EPPKALECANPAVLQGIRGNIQETLTQEARSFAREDGRQF

```

-392-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf25a.pep	VDADXI	IAAAXXXS	LEHASETQ	EGGRTFC	XADLNITV	PSETLADAKANSPLLYGETAL
	orf25-1	VDADKI	IAAAYGLA	FSLEHASETQ	EGGRTFC	IADLNITV	PSETLADAKANSPLLYGETAL
		70	80	90	100	110	120
10	orf25a.pep	SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQXAFVDNTVGMAAQTL	SAALLPYGVKSIV				
	orf25-1	SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQTAFVDNTVGMAAQTL	SAALLPYGVKSIV				
		130	140	150	160	170	180
15	orf25a.pep	MIDGKAVKKEDAVRIXSXXAREXEPSKXXPEDILEHNAAGGDADVPQAGEDAPEPEILHP					
	orf25-1	MIDGKAVKKEDAVRILSGKAREEPEPSKPTPEDILEHNAAGGDAGVPQAAGAPEPEILHP					
		190	200	210	220	230	240
20	orf25a.pep	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLD	TDVQKELVGE	EXRKWAQEKISNC			
	orf25-1	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLD	TDVQKELVGE	QKRWAEKISNC			
		250	260	270	280	290	300
25	orf25a.pep	RQAAAQADRQ	EYAEYLKLC	QDTRMTRERIQYLRGYSID			
	orf25-1	RQAAAQADRQ	EYAEYLKLC	QDTRMTRERIQYLRGYSID			
		310	320	330	339		
30	orf25a.pep						
	orf25-1						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from

35 *N.gonorrhoeae*:

	orf25.pep	TDVQKELVGEQRKWAQEKISNCRQAAAQAD	30
	orf25ng	VTVSRGEVEEARVQNQRAESEITKLWGGLD	TDVQKELVGEQRKWAQEKISNCRQAAAQAD 308
40	orf25.pep	RQEYAEYLKLCQDTRMTRERIQYLRGYSID	60
	orf25ng	RQEYAEYLKLCQDTRMTRERIQYLRGYSID	338

The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

45	1	ATGTATCGGA	AACTCATTCG	GCTGCCGTTT	GCCCTGCTGC	TGCGAGCGTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCGTTGGA	ATGCGCCAAC	CCCGCCGTGT
	101	TGCAGGACAT	ACGCGGCAGT	ATTGAGGAAA	CGCTCACGCA	GGAAGCCGCT
	151	TCTTTCGCGC	GCGAAGACGG	CAGGCAGTTT	GTCGATGCCG	ACAAAATTAT
	201	CGCCGCCGCC	TACGGTTTGG	CGTTTTCTTT	GGAACACGCT	TCGGAAACGC
50	251	AGGAAGGCGG	GCGCACGTTT	TGTATCGCCG	ATTGAACAT	TACCGTGCCG
	301	TCTGAAACGC	TTGCCGATGC	CGAGGCAAAC	AGCCCCCTGC	TGTATGGGGA
	351	AACGTCTTTG	GCAGACATCG	TGCAGCAGAA	GACGGGCGGC	AATGTCGAGT
	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCCTGCC	CGCCAAAGAC
	451	GCTCGGACGG	CATTTATCGA	CAACACGGTC	GGTATGGCGA	CGCAAACGCT
55	501	GTCTGCCGCG	TTGCTGCCCT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
	551	GCAAGGCGGT	GACAAAAGAA	GACGCGGTCA	GGGTTTGTAG	CGGCAAAGCC
	601	CGTGAAGAAG	AACCGTCCAA	ACCCACCCCC	GAAGACATT	TGGAACACAA
	651	TGCCGCCGGC	GGCGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGCACCCG
	701	AACCCGAAAT	CCTGCATCCC	GACGACGTCG	AGCGTGCCGA	TACCGTTACC
	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAACC	AACGTGCGGA
60	801	ATCCGAAATT	ACCAAACTTT	GGGAGGACT	CGATACCGAC	GTGCAAAAAG
	851	AGTTGGTCCG	CGAACAGCGC	AAGTGGGCGC	AGGAAAAAAT	CAGcaactgc
	901	cgACAAGCCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
	951	GCTCCAATGC	GACACGCGGA	TGACGCGCGA	ACggaTACAG	TATCTTCGCG
	1001	GCTATTCCAT	CGATTAG			

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

-393-

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQDIRGS IQETLTQEAR  
 51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP  
 101 SETLADAEAN SPLLYGETSL ADIVQKQTGG NVEFKDGVLT AAVRFLPAKD  
 151 ARTAFIDNTV GMATQTLCAA LLPYGVKSIV MIDGKAVTKE DAVRVLSGKA  
 201 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDVERADTVT  
 251 VSRGEVEEAR VQNQRAESEI TKLWGGGLTD VQKELVGEQR KWAQEKISNC  
 301 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID\*

ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

10	orf25-1.pep	10	20	30	40	50	60
	orf25ng	10	20	30	40	50	60
15	orf25-1.pep	70	80	90	100	110	120
	orf25ng	70	80	90	100	110	120
20	orf25-1.pep	130	140	150	160	170	180
	orf25ng	130	140	150	160	170	180
25	orf25-1.pep	190	200	210	220	230	240
	orf25ng	190	200	210	220	230	240
30	orf25-1.pep	250	260	270	280	290	300
	orf25ng	250	260	270	280	290	300
35	orf25-1.pep	310	320	330	339		
	orf25ng	310	320	330	339		

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

50 ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and
   
 55 that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

### Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

5      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
      51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
     101  GCATCGGTAT TCTGGwysGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
     151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
     201  CGsyGATTGG TCGCTGGGCA AACC AAAAAT CTTGGTTTTT CkGATACTTT
     251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA T.....

10      //
      851  .....AC TTCGCTGGTA
      901  TTCGGCGGCA CTTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGCTCGGCAC
      951  GATTAAACC GCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
     1001  TGTTCCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
     151  1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
     1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
     1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
     1201  ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
     1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
     20  1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
     1351  GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
     1401  CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
     1451  TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
     1501  AAAAAA..

```

25 This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

      1  MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILXX VAFVLGGNPV
      51  DGLTHLKDMV VGLAWSDXDW SLGKPKILVF XILLGIFTSI LTYSGSN...

30      //
     251  .....TSLV
     301  FGGTCGVFAV VLCTLTGTTK ADYPKAVWQG AKSMFGAIAI LILAWLISTV
     351  VGEMHTGDYL STLVAAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
     401  IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNCHI
     451  DHVTSQLPYA LTVAAAAASC YLALGLTKSA LLGFGTGIV LAVLIFLLKD
     501  KK..

```

35 Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
      51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
     101  GCATCGGTAT TCTGGTCGGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
     151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
     201  CCGCGATTGG TCGCTGGGCA AACC AAAAAT CTTGGTTTTT CTGATACTTT
     251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
     301  GCGGACTGGG CAAAACGGCA CATTAAAAAC CCGCGCGGCG CGAAAAATGCT
     351  GACCGCCTGC CTCGTGTTTC TAACCTTTAT CGACGACTAT TTCCACAGTC
     401  TCGCCGTCCG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
     451  CGCACCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCTC CTATGTGCGT
     501  GCTGATGCCC GTTTCAAGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
     551  GACTGCTCGT TACCTACAAA ATCACC GAAT ACACGCCGAT GGGGACGTTT
     601  GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
     651  GTTCGTCGTC GCATGTTTCT CCTTCGACAT CGGCTCGATG GCACGTTTCG
     701  AACAAGCCGC GTTGAACGAA GCCACGATG AACTGCCGT TTCAGACGCT
     751  ACCAAAGGTC GTGTTTACGC ACTGATTATT CCGTTTTTGG CCTTAATCGC
     801  CTCACCGGTT TCCGCCATGA TCTACACCGG CCGCGAGGCA AGCGAAACCT
     851  TCAGCATTTT GGGGGCATTT GAAAACACGG ACGTAAACAC TTCGCTGGTA
     901  TTCGGCGGCA CTTGCGGCGT CCTTGCCGTC GTTCTCTGCA CGCTCGGCAC
     951  GATTAAACC GCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
    1001  TGTTCCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
    1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
    1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
    1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
    1201  ATTGCCGCCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
    1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
    1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC

```

	1	<u>MQLIDYSHSF</u>	<u>FSVVPFFLAL</u>	<u>ALAVITRRVL</u>	<u>LSLGIGILVG</u>	<u>VAFLVGGNPV</u>
	51	<u>DGLTHLKMV</u>	<u>VGLAWSGDW</u>	<u>SLGKPKILVF</u>	<u>LILLGIFTSL</u>	<u>LTYSGSNQAF</u>
	101	<u>ADWAKRHKN</u>	<u>RGAKMLTAC</u>	<u>LVEVTFIDDY</u>	<u>FHSLAVGAIA</u>	<u>RPVTDKFKVS</u>
10	151	<u>RTKLAYILDS</u>	<u>TAAPMCVLMP</u>	<u>VSSWGASIIA</u>	<u>TLAGLLVTYK</u>	<u>ITEYTPMGTF</u>
	201	<u>VAMSLMNYIA</u>	<u>LFALIMVFVV</u>	<u>AWFSFDIGSM</u>	<u>ARFEQAALNE</u>	<u>AHDETAVSDA</u>
	251	<u>TKGRVYALII</u>	<u>PVLALIASTV</u>	<u>SAMIYTGAAQ</u>	<u>SETFSILGAF</u>	<u>ENTDVNTSLV</u>
	301	<u>FGGTCGVLA</u>	<u>VLCTLTGTIK</u>	<u>ADYPKAVWQG</u>	<u>AKSMFEGAI</u>	<u>LILALISTIS</u>
	351	<u>VGENHTGDYL</u>	<u>STLVAGNIHP</u>	<u>GFLPVILFLL</u>	<u>ASVMAFATGT</u>	<u>SWGTFGLMLP</u>
	401	<u>IAAAMAVKVE</u>	<u>PALIIPCMSA</u>	<u>VMAGAVCGDH</u>	<u>CSPISDITIL</u>	<u>SSTGARCNIH</u>
15	451	<u>DHVTSQLPYA</u>	<u>LTVA AAAASG</u>	<u>YLALGLTKSA</u>	<u>LLGFGTGTIV</u>	<u>LAVLIFLLKD</u>
	501	<u>KKRANA*</u>				

```

50                                10      20      30      40      50      60
   orf26.pep  MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV
               |||||
   orf26a     MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
               |||||

55                                10      20      30      40      50      60
   orf26.pep  VGLAWSDXDWSLGKPKILVFXILLGIFTSLLTSGSNXX-----
               |||||
   orf26a     VGLAWSGDGWSLGKPKXLVFLILLGIFTSLLTSGSNQAFADWAKRHIKNRRGAKMLTAC
               |||||
60                                70      80      90     100     110     120

```

45	1	ATGCAGCTGA	TCGACTATTTC	ACATTTCATTT	TTCTCGGTTG	TGCCACCCCTT
	51	TTTGGCACTG	GCACTTGCCG	TCATTACCCG	CCGGCTACTG	CTGTCTTTAG
	101	GCATCGGTAT	TCTGGTCGGC	GTTGCCTTTT	TGGTCGGCCG	CAACCCCGTC
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	CTCGGCTTGG	CTTTGGTCAGA
50	201	CGGCGATTGG	TCGCTGGGCA	AACCAAAANT	CTTGGTTTTTC	CTGATACTTT
	251	TGGGTATTTT	TACTTCCTCTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGTTT
	301	GCCGACTGGG	CAAAACGGCA	CATTA AAAAC	CGGCGCGCTG	CGAAAATGCT
	351	GACCGCCTGC	CTCGTGTTCG	TAACCTTTAT	CGACGACTAT	TTCCACAGTC
55	401	TCGCCGTCGG	TGCGNTTGGC	CGCCCCGTTA	CCGACAAGTT	TAAAGTTTCC
	451	CGCGCCAAAC	TCGCCCTACAT	CCTCGACTCC	ACTGCCCGCG	CTATGTGCGT
	501	GCTGATGCCC	GTTTCAAGCT	GGGGCGCGTC	GATTATCGCC	ACGCTTGCCC
	551	GACTGCTCGT	TACCTACAAA	ATCACCGAAT	ACACGCCGAT	GGGGACGTTT
60	601	GTCGCCATGA	GCCTGATGAA	CTATTACGCA	CTGTTTGCCC	TGATTATGGT
	651	GTTTGTCTGT	GCATGGTTCT	CCTTCGACAT	CGGCTCGATG	GCACGTTTCG
	701	AACAAGCCGC	GTTGAACGAA	GCCCACGATG	AAACTGCCGT	TTCAGACGGC
	751	AGCTGGGGCA	GGGTTTACGC	ATTGATTATT	CCCGTTTTGG	CCTTAAATCGC
65	801	CTCAACGGTT	TCCGCCATGA	TCTACACCGG	TGCACAGGCA	AGCGAAACCT
	851	TCAGCATTTT	GGGTGCATTT	GAAAAATACGG	ACGTGAACAC	TTTCGTGGTA
	901	TTCCGCGGCA	CTTGC GCGT	GCTTGCCGTC	GCTCTCTGCA	CGCTCGGCGT
	951	GATTAAAATC	GCCGATTATC	CCAAAGCCGT	TTGGCAGGGT	GCGAAATCCA
70	1001	TGTTCGGCGC	AATCGCCATT	TTAATCCTTG	CCTGGCTCAT	CAGTACGGTT
	1051	GTCGGCGAAA	TGCACACAGG	CGACTACCTC	TCCACGCTGG	TTGCGGGCAA
	1101	CATCCATCCC	GGCTTCTGN	CCGTTCATCCT	TTTTCTGCTC	GCCACGCTGA
	1151	TGGCGTTTGC	CACAGGCACA	AGCTGGGGGA	CGTTCCGGCAT	CATGCTGCCG
75	1201	ATTGCCGCGG	CCATGGCGGT	CAAAGTCGAT	CCCTCACTGA	TTATCCCGTG
	1251	TATGTCGGCC	GATGATGGCG	GGGCGGTATG	CGGCGACAC	TGCTCGCCCA
	1301	TTTCCGACAC	GACCATTCTG	TCGTCCACCG	CGGCGCGCTG	CACACCATTA



-397-

1351 GACCACGTTA CNTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC  
 1401 CGCATCGGGN TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGTT  
 1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAGAT  
 1501 AAAAAACGCG CCAACGCCTG A

5 This encodes a protein having amino acid sequence <SEQ ID 694>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLVGNPV  
 51 DGLTHLKDMV VGLAWSDDGW SLGKPKXLVF LILLGIFTSL LTYSGSNQAF  
 101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHSLAVGAXA RPYTDKFKVS  
 151 RAKLAYILDS TAAPMCVLMF VSSWGASIIA TLAGLLVYK ITEYTPMGTF  
 10 201 VAMSLMNYA LFALIMVFWV AWFSFDIGSM ARFEQAALNE AHDETAVSDG  
 251 SWGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTVDNTSLV  
 301 FGGTCGVLAV VLCTLTGTIKI ADYPAVWQG AKSMFGAIAI LILAWLISTV  
 351 VGENHTGDYL STLAVAGNIHP GFLXVILFLL ASVMAFATGT SWGTFGIMLP  
 401 TAAAMAVKVD PSLIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI  
 15 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGXTGIV LAVLIFLLKD  
 501 KKRANA\*

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

20	orf26a.pep	10	20	30	40	50	60
	orf26-1	10	20	30	40	50	60
25	orf26a.pep	70	80	90	100	110	120
	orf26-1	70	80	90	100	110	120
30	orf26a.pep	130	140	150	160	170	180
	orf26-1	130	140	150	160	170	180
35	orf26a.pep	190	200	210	220	230	240
	orf26-1	190	200	210	220	230	240
40	orf26a.pep	250	260	270	280	290	300
	orf26-1	250	260	270	280	290	300
50	orf26a.pep	310	320	330	340	350	360
	orf26-1	310	320	330	340	350	360
55	orf26a.pep	370	380	390	400	410	420
	orf26-1	370	380	390	400	410	420
60	orf26a.pep	430	440	450	460	470	480
	orf26-1	430	440	450	460	470	480
65	orf26a.pep	490	500				
	orf26a.pep	LLGFGXTGIVLAVLIFLLKDKKRANAX					

-398-

```

      |||||:|||||
orf26-1  LLGFGTTGIVLAVLIFLLKDKKRANAX
              490       500

```

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

```

10 orf26.pep  MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV 60
    orf26ng   MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFVGGNPVDGLTHLKDMV 60

15 orf26.pep  VGLAWSDXDWSLGGPKILVFXILLGIFTSLLTYSGSN 97
    orf26ng   VGLAWADGDWSLGGPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC 120

//

20 orf26.pep  TSLVFGGTCGVFAVVLCTLGTIKTADYPKA 326
    orf26ng   ASTVSAMIYTGQAASETFSILGAFENTDVNTSLVFGGTCGVLA VVLCTFGTIKTADYPKA 326

25 orf26.pep  VWQGA KSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLLASVMAF 386
    orf26ng   VWQGA KSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLLASVMAF 386

30 orf26.pep  ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPI SDTTILSSTGAR 446
    orf26ng   ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPI SDTTILSSTGAR 446

30 orf26.pep  CNHIDHVTSQLPYALTVA AAAAAAGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK 502
    orf26ng   CNHIDHVTSQLPYALTVA AAAAAAGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK RADV 506

```

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

```

35 1 ATGCAGCTGA TTGACTATTC ACATTCATT TTTCTCGGTTG TGCCACCCTT
   51 TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
  101 GCATCGGTAT TTTGGTCGGC GTTGCCCTTT TGGTCGGCGG CAACCCCGTC
  151 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTGGGCAGA
  201 CGGCGATTGG TCGCTGGGCA AACCAAAAT CTTGGTTTTT CTGATACTTT
  251 TGGGCACTTT CACTTCACTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
  301 GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGTGCGGCG CGAAAATGCT
  351 GACCGCCTGC CTCGTGTTTC TAACCTTTAT CGACGACTAT TTCCACAGCC
  401 TCGCCGTCGG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
  451 CGCGCCAAAC TCGCCTACAT CCTCGACTCC ACTGCCTCGC CCATGTGCGT
  501 GCTGATGCCC GTTTCAGGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
  45 551 GATTGCTCGT TACCTACAAA ATTACCGAAT ACACGCCGAT GGGGACGTTT
  601 CTCGCCATGA GCCTGATGAA CTATTACGCG CTGTTTGCCC TGATTATGGT
  651 ATTCGTCGTC GCATGGTTCT CCTTCGACAT CGGCTCGAtg gCGCGTTTCG
  701 AACAGGCTGC GTTGAACGAA gccccaggacg aaaccgccgc tTCAGACgCT
  751 ACCAAAGGTC GTGTTTACGC ATTGATTAT CCCGTTTGG CCTTAATCGC
  801 CTCACCGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT
  851 TCAGCATTTT GGGGGCATTT GAAAATACCG ACGTAAACAC TTCGCTGGTA
  901 TTCGGCGGCA CTGCGGCGCT GCTTGCCGTC GTCCTCTGCA CGTTCGGCAC
  951 GATTAAAAAC GCCGATTATC CCAAAGCCGT GTGGCAGGGT GCGAAATCCA
  1001 TGTTCGGCGC AATCGCCATT TTAATCCTCG CCTGGCTCAT CAGTACGGTT
  55 1051 GTCGGCGAAA TGCACACGGG CCACTACCTC TCCACGCTGG TTGCGGGCAA
  1101 CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
  1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
  1201 ATTGCCGCGC CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTAAtcccGTG
  1251 TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGTTCGCCCA
  1301 TCTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
  1351 GACCACGTTA CCTCGCAACT GCCTTATGCC CTGACGGTTG CCGCCGCGCG
  1401 CGCATCGGCG TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGCGT
  1451 TTGGCACGAC CGGTATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
  1501 AAAAAACGCG CCGACGTTT A

```

65 This encodes a protein having amino acid sequence <SEQ ID 696>:

10

15

20

25

30

35

40

45

50

55

65

In addition, ORF26 ng shows significant homology to a hypothetical *H.influenzae* protein:

-400-

sp|P44263|YF86\_HAEIN HYPOTHETICAL PROTEIN HI1586 >gi|1074850|pir||C64037  
 hypothetical  
 protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.  
 influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519  
 Score = 538 bits (1370), Expect = e-152  
 Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

5 Query: 1 MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXAFLVGGNPVDGLTHLKDMV 60  
 M+LID+S S +S+VP LA+ LA+ TRR L +L V  
 10 Sbjct: 14 MELIDFSSSVWSIVPALLAILAIATRRVLVSLSAGIIGSLMLSDWQIGSAFNVLVKNV 73

Query: 61 VGLAWADGDWSL GKPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHKRNRCGAKMLTAC 120  
 V L +ADG+ + I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A  
 15 Sbjct: 74 VSLVYADGEIN-SNMNIVFLLLLLGVLTALLTVSGSNRAFAEWAQSRIGRRGAKLLAAS 132

Query: 121 LVFVTFIDDYFHSLAVGAIARPVTDKFKVSRAKLAYILDSTASPMCVLMPVSSWGASIIA 180  
 LVFVTFIDDYFHSLAVGAIARPVTD+FKVSRAKLAYILDSTA+PMCV+MPVSSWGA II  
 Sbjct: 133 LVFVTFIDDYFHSLAVGAIARPVTDKFKVSRAKLAYILDSTAAPMCVMMPVSSWGAYIIT 192

20 Query: 181 TLAGLLVYTKITEYTPMGTFVAMSLMNYIALFALIMVFWVWAFSFDIGSMARFEQAALNE 240  
 + GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+FSFDI SM R E+ AL  
 Sbjct: 193 LIGLLATYSITEYTPIGAFVAMSSMNFYAIFSIIMVFFVAYFSFDIASMRHEKLALKN 252

25 Query: 241 AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQA----SETFSILGAFENTDVN 296  
 +D+ TKG+V LI+P+L LI +TVS MIYTGA+A + FS+LG FENT V  
 Sbjct: 253 TEDQLEEETGTGQVRNLILPILVLIATVSMMIYTGAELAADGKVFSVLGTFENTVVG 312

30 Query: 297 TSLVFGGTCGVL--AVVLCTFGTIKTADYPKAVWQAKSMFGXXXXXXXXXXXXSTVVGEM 354  
 TSLV GG C ++ +++ + +Y ++ G KSM G + +VG+M  
 Sbjct: 313 TSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGAIAILFFAWTINKIVGDM 372

35 Query: 355 HTGDYLSLTVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPFAAAMAVKVEPALI 414  
 TG YLS+LV+GNI FLPVILF+L + MAF+TGTSGWGTFGIMLPFAAAMA P L+  
 Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLPFAAAMAANAPELL 432

Query: 415 IPCMSAVMAGAVCGDHCSPISTTILSSTGARC�HIDHVT SQXXXXXXXXXXXXXXXXXXXX 474  
 +PC+SAVMAGAVCGDHCSPISTTILSSTGA+CNHIDHVT+Q  
 Sbjct: 433 LPCLSAVMAGAVCGDHCSPISTTILSSTGAKCNHIDHVTQLPYAATVATATSIGYIVV 492

40 Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501  
 S L GF T + L V+IF +K +  
 Sbjct: 493 GFTYSGLAGFAATAVSLIVIIIFAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*,  
 45 and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

50 1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA  
 51 CGATGAGCCT GCCAAAATTC TGA CTGGGA TGAAAGCGGC CGATTACTCT  
 101 CGGAACTGTC TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG  
 151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTATATCAGG ATGACAAGTT  
 201 GGTGAGGAAA ACCCAGTGGG ATAAGGATGG TTATTATATC GAACCTGA

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

55 1 ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HQNGVVLEW  
 51 YEDGSKKSEX VYQDDKLVRK TQWDKDGILI EP\*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

60 1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC  
 51 GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGAA  
 101 AGCTGACGGC GACGATGCTT TCTGCCGCTT ATATCAGGCA ATATAGTGTG  
 151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA

10

15

20 ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N. meningitidis*:

30

35

40

45

50

55 ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

orf27a.pep MKKLSRIVFSTVLLGFS AALPAQXYSVYFNONGKLTATXSSAAYIROQSV AEGIAHQAQXF

-402-

	orf27-1	 MKKLSRIVFSTVLLGFS AALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF 10 20 30 40 50 60
5	orf27a.pep	70 80 90 100 110 120 XYP SMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFXGQKKMAGGFSKGKPDGEWVNWYP 
10	orf27-1	70 80 90 100 110 120 YYP SMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGKPDGEWVNWYP 
	orf27a.pep	130 140 150 160 170 180 NGKKS AVMPYKNGLSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVN 
15	orf27-1	130 140 150 160 170 180 NGKKS AVMPYKNGLSEGTGYRYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN 
	orf27a.pep	190 200 210 220 230 240 DEPAKILTWDSEGRLLSELSIRHHXRNQGVVLEWYEDGSKKXEA VYQDDKLVRKTQWDKDG 
20	orf27-1	190 200 210 220 230 240 DEPAKILTWDSEGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG 
25	orf27a.pep	YLIEPX 
	orf27-1	YLIEPX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from  
30 *N.gonorrhoeae*:

	orf27.pep	KQWYADXSIKTEMVMVNDEPAKILTWDSESG 30 
	orf27ng	LSEGTGYRYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVNDEPAKILTWDSESG 193
35	orf27.pep	RLLSELSIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDG YLIEP 82 
	orf27ng	RLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG YLIEP 245

The complete length ORF27ng nucleotide sequence <SEQ ID 703> is:

40	1	ATGAAGAAAT	TATCTCGGAT	TGTATTTTCA	ATCGTACTGT	TGGGTTTTTC
	51	GGCCGCTTTG	CCGGCGCAGA	CCTATTCTGT	TTATTTTAAT	CAGAACGGGA
	101	AACTGACGGC	GACGATGTCT	TCTGCCGCTT	ATATCAGGCA	ATATAGTGTG
	151	GCGGCGGGTA	TGCGACACGC	GCAGGATTTT	TATTATCCGT	CGATGAAGAA
	201	ATATTCCGAA	CCTTATATCG	TTGCTTCAAC	GCAAATCAAA	TCTTTTGTGC
45	251	CTACCCTGCA	AAACGGTATG	TTGATTTTGT	GGCATTTTAA	TGGTCAGAAA
	301	AAAATGGCGG	GGGGCTTCAG	CAAGGGTAAG	CCGGACGGGG	AATGGGTCAA
	351	CTGGTATCCG	AACGGTAAAA	AATCTGCGGT	TATGCCTTAT	AAAAATGGCT
	401	TGAGTGAGGG	TACGGGATAC	CGTTATTACC	GTAACGGCGG	CAAGGAAAGC
	451	GAAATCCAGT	TTAAGCAAAA	TAAGGCGAAC	GGCGTATGGA	AGCAATGGTA
50	501	TGCCGATGGA	AGTATCAAGA	CGGAAATGGT	TATGGTCAAC	GATGAGCCTG
	551	CCAAAATTCT	GACTTGGGAT	GAAAGCGGCC	GATTACTTTC	GGAAGTGTCT
	601	ATCCGCCACC	ATAAACGCAA	CGGGGTGGTT	TTGGAGTGGT	ATGAAGATGG
	651	TTCTAAAAAG	AGCGAGGCTG	TTATCAGGA	TGACAAGTTG	GTCAGGAAAA
	701	CCCAATGGGA	TAAGGATGGT	TATTTAATCG	AACCCTGA	

This encodes a protein having amino acid sequence <SEQ ID 704>:

55	1	MKKLSRIVFS	IVLLGFS AAL	PAQTYSVYFN	QNGKLTATMS	SAAYIRQYSV
	51	AAGIAHAQDF	YYP SMKKYSE	PYIVASTQIK	SFVPTLQNGM	LILWHFNGQK
	101	KMAGGFSKGK	PDGEWVNWYP	NGKKS AVMPY	KNGLSEGTGY	RYRNRNGKES
	151	EIQFKQNKAN	GVWKQWYADG	SIKTEMVMVN	DEPAKILTWD	ESGRLLSELS
60	201	IRHHRNGV	LEWYEDGSKK	SEAVYQDDKL	VRKTQWDKDG	YLIEP*

ORF27ng and ORF27-1 show 98.8% identity in 245 aa overlap:

	orf27-1.pep	10 20 30 40 50 60 MKKLSRIVFSTVLLGFS AALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF
--	-------------	--

10

15

20

25

30

35

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

45

50

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHX LSGFYWHAHE  
51 MWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTIFWL AARIAAFIPG  
101 WGASASGILG TLEFFYGAVC MALPVIRSON ORNYVAVEAL FVLGGTHAAF

151 HVQLHNGNLG GLLSGLQSGL VM

Further work revealed the complete nucleotide sequence <SEQ ID 707>:

```

      1 ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
      51 TTCGCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151 ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201 CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGGCGGC GTTCTGGTCG
251 GCTTGACTAT CTTTGGCTG GCTGCGCGGA TTGCCGCTT TATCCCGGT
301 TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTT TCTGGTACGG
10 351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAA CAACGCAACT
401 ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGCACGCA TGCGGCGTTC
451 CACGTCCAGC TGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
501 TGGGGCTTG GTGATGGTGT CGGGTTTAT CGGTCTGATT GGTACGCGGA
15 551 TTATTTCGTT TTTTACGTCC AAACGCTTGA ATGTGCCGCA GATTCCCAGT
601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGA CTGCCAT
651 GCTGATGGCG CACGGTGTGT TGGCTTGGCT GTCTGCCGTT TTTGCCTTG
701 CGGCAGGTGT GATTTTTACC GTGCAGTGT ACCGCTGTTG GTATAAACCC
751 GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCGGCTTTCC
20 851 TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901 TGGGGCATGA TGGCGCGTAC CGCGCTGGT CATACGGGCA ATCCGATTTA
951 TCCGCCGCCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CCGCGTCCG TATGTTTGGC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATGC
25 1101 GTGGAAGTAT ATTCCTTGGC TGATTCTGCC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA

```

This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

```

      1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
      51 MIWGYAGLVV IAFLLTAVAT WTGQPTRGG VLVGLTIFWL AARIAAFIPG
101 WGASASGILG TLFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSGL VMVSGFIGLI GTRIIISFFTS KRLNVPQIPS
201 PKWVAQASLW LPMLTAMLMA HGVLAWLSAV FAFAAGVIFT VQVYRWYKPK
251 VLKEPMLWIL FAGYLFGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT
301 LGM MARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
35 351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GREPG*

```

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*

40 *meningitidis*:

```

      10      20      30      40      50      60
orf47.pep MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFYGTGTHXLSGFYWHAHEMIWGYAGLVV
      |||
orf47a    MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60

      70      80      90      100     110     120
orf47.pep IAFLLTAVATWTGQPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      |||
orf47a    IAFLLTAVATWTGQPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      70      80      90      100     110     120

      130     140     150     160     170
orf47.pep MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVM
      |||
orf47a    MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVSGFIGLI
      130     140     150     160     170     180

orf47a    GTRIIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLWLSAFAFAAGVIFT
      190     200     210     220     230     240

```



The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

```

1  ATGAAATTTA CCAAGCACCC CGTTTGGGCA ATGGCGTTCC GCCCGTTTTA
51  TTCCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151 ATGATTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201 CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGCGGC GTTCTGGTCG
251 GCTTGACTAT CTTTGTGCTG GCTGCGCGGA TTGCCGCTT TATCCCGGT
301 TGGGGTGCCT CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGGTACGG
351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAATT
401 ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGTACGCA CGCGGCGTTC
451 CACGTCCAGC TGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
501 GTCGGGCTTG GTGATGGTGT CCGGTTTTAT CCGTCTGATT GGTACGCGGA
551 TTATTTCGTT TTTTACGTCC AAACGGTTGA ATGTGCCGCA GATTCCAGT
601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGACCGCCAT
651 CTGTATGGCG CACGGCGTGA TGCCTTGGCT GTCGCGGCT TTCGCGTTG
701 CGGCAGGTGT GATTTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAGCCT
751 GTGTTGAAAG AGCCGATGCT GTGATTCTG TTTGCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCCGCTTTCC
851 TCAATCTGGG TGTGCACTG ATCGGGTGC GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CACACGGGCA ATCCGATTTA
951 TCCGCCGCCC AAAGCCGTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CCGCCGTCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATACGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTT TGGTGTATGC
1101 GTGGAAGTAT ATTCTTGGC TGATTCTGTC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 710>:

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
51  MIWGYAGLVV IAFLLTAVAT WTGPPTRGV VLVGLTIFWL AARIAAFIPG
101 WGASASGILG TLFFWYGAVC MALPVIRSQN QRNYVAVFAL FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GTRIIISFTS KRLNVPQIPS
201 PKWVAQASLW LPMLTAMLMH HGVMPWLSAA FAFAAGVIFT VQVYRWYKYP
251 VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFLNLGVHL IGVGIGVILT
301 LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

```

10      20      30      40      50      60
orf47a.pep  MKFTKHPVWMAFAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
40      10      20      30      40      50      60
orf47-1     MKFTKHPVWMAFAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV

70      80      90      100     110     120
orf47a.pep  IAFLLTAVATWTGPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
45      10      20      30      40      50      60
orf47-1     IAFLLTAVATWTGPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC

130     140     150     160     170     180
orf47a.pep  MALPVIRSQNQRNYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSGLMVSGFIGLI
50      10      20      30      40      50      60
orf47-1     MALPVIRSQNQRNYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSGLMVSGFIGLI

190     200     210     220     230     240
orf47a.pep  GTRIIISFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMHGVMPWLSAAFAFAAGVIFT
55      10      20      30      40      50      60
orf47-1     GTRIIISFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMHGVMPWLSAAFAFAAGVIFT

250     260     270     280     290     300
orf47a.pep  VQVYRWYKYPVLKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVILT
60      10      20      30      40      50      60
orf47-1     VQVYRWYKYPVLKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVILT

310     320     330     340     350     360

```

orf47a.pep	LGMMARTALGHTGNPIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
orf47-1	LGMMARTALGHTGNPIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
	310 320 330 340 350 360
orf47a.pep	370 380
orf47-1	LALLVYAWKYIPWLIRPRSDGRPGX
	370 380

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

15	ORF47	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWAHHEMIWGYAGLVV	60
	ORF47ng	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWAHHEMIWGYAGLVV	60
20	ORF47	IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC	120
	ORF47ng	IAFLLTAVATWTGQPPTRGGLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC	120
25	ORF47	MALPVIRSQNRNYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSGGLVM	172
	ORF47ng	MALPVIRSQNRNYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSGGLVMVWGFGLI	180

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

	1	<u>MKFTKHPVWA</u>	<u>MAFRPFYSLA</u>	<u>ALYGALSVLL</u>	<u>WGFGYGTGTHE</u>	<u>LSGFYWAHA</u>
	51	<u>MIWGYAGLVV</u>	<u>IAFLLTAVAT</u>	<u>WTGQPPTRG</u>	<u>VLVGLTAFWL</u>	<u>AARIAAFIPG</u>
30	101	<u>WGAAASGILG</u>	<u>TLFFWYGAVC</u>	<u>MALPVIRSQN</u>	<u>RRNYVAVFAI</u>	<u>FVLGGTHAAF</u>
	151	<u>HVQLHNGNLG</u>	<u>GLLSGLQSG</u>	<u>VMVWGFGLI</u>	<u>GMKIISFFTS</u>	<u>KRLKLPQIPS</u>
	201	<u>PKWVAHASLW</u>	<u>LPMLNAILMA</u>	<u>HRVMPWLSAA</u>	<u>FFFAAGVIFT</u>	<u>VQVYAGGITP</u>
	251	<u>IEETSCGSVA</u>	<u>GICYRLGNSS</u>	<u>G</u>		

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

	TM segments in ORF47ng			
40	INTEGRAL	Likelihood = -5.63	Transmembrane	52 - 68
	INTEGRAL	Likelihood = -3.88	Transmembrane	169 - 185
	INTEGRAL	Likelihood = -3.08	Transmembrane	82 - 98
	INTEGRAL	Likelihood = -1.91	Transmembrane	134 - 150
	INTEGRAL	Likelihood = -1.44	Transmembrane	107 - 123
	INTEGRAL	Likelihood = -1.38	Transmembrane	227 - 243

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

45	1	ATGAAATTTA	CCAAACATCC	CGTCTGGGCA	ATGGCGTTCC	GCCCGTTTTA
	51	TTCACCTGGCG	GCACTGTACG	GCGCATGTGC	CGTATTGCTG	TGGGGTTTCG
	101	GCTACACGGG	AACGCACGAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAG
	151	ATGATTTGGG	GTTATGCCGG	TCTCGTCGTC	ATCGCCTTCC	TGCTGACCGC
50	201	CGTCGCCACT	TGGACGGGAC	AGCCGCCAC	GAGGGCGGC	GTTCTGGTCG
	251	GCTTGACCGC	CTTTTGGCTG	GCTGCGCGGA	TTGCCGCCTT	TATCCCGGGT
	301	TGGGGTGCGG	CGGCAAGCGG	CATACTCGGT	ACGCTGTTT	TCTGGTACGG
	351	CGCGGTGTGC	ATGGCTTTGC	CCGTATCCG	TtcgCAAAAC	CGGCGCAACT
	401	ATGtcgCCGT	ATTGCAATA	TTTGTGCTGG	GCGGTACGCA	TGCGgcgTTC
	451	CACGtccAgc	tGCACAACGG	CAACCTAGGC	GGACTCTTGA	GCGGATTGCA
55	501	GTCGGGCGCTG	GTTATGGTGT	CGGGCTTTAT	CGGCCTGATT	GGGATGAGGA
	551	TTATTTCGTT	TTTACGTCC	AAACGGTTGA	ACGTGCCGCA	GATTCCCAGT
	601	CCGAAATGGG	TGGCGCAGGC	TTGCTGTGG	CTACCCATGC	TGACCGCCAT

-407-

```

5  651 ACTGATGGCG CACGGCGTGA TGCCCTGGCT GTCGGCGGCT TTCGCGTTTG
    701 CGGCGGGCGT GATTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
    751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
    801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCTGCCTTCC
    851 TCAATCTGGG CGTACATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
    901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CACACGGGCA ATTCGATTTA
    951 TCCGCCGCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
10 1001 CCGCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
    1051 AGCATCCGCA CGTCTTCGGT TTGTTTGCA CTCGCGCTGC TGGTGTATGC
    1101 GTGGAAATAC ATTCCGTGGC TGATCCGTCC GCGTTCGGAC GGCAGGCCCG
    1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

```

15  1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
    51  MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTAFWL AARIAAFIPG
    101  WGAAASGILG TLFFWYGAVC MALPVIRSON RRNYVAVFAI FVLGGTHAAF
    151  HVQLHNGNLG GLLSGLQSGL VMVSGFIGLI GMRIISFFTS KRLNVPQIPS
    201  PKWVAQASLW LEMLTAILMA HGVMPWLSAA FAFAGVIFT VQVYRWYKYP
    251  VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT
    301  LGMMARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
20  351  SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

```

25  orf47-1.pep  10      20      30      40      50      60
    MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
    orf47ng-1    10      20      30      40      50      60
    MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV

30  orf47-1.pep  70      80      90      100     110     120
    IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
    orf47ng-1    70      80      90      100     110     120
    IAFLLTAVATWTGQPPTRGVVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC

35  orf47-1.pep  130     140     150     160     170     180
    MALPVIRSONQRNYVAVFALEVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVMVSGFIGLI
    orf47ng-1    130     140     150     160     170     180
    MALPVIRSONRRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVMVSGFIGLI

40  orf47-1.pep  190     200     210     220     230     240
    GTRIIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAGVLAWSAVFAFAAGVIFT
    orf47ng-1    190     200     210     220     230     240
    GMRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAILMAHGVMPWLSAAFAFAAGVIFT

45  orf47-1.pep  250     260     270     280     290     300
    VQVYRWYKYPVLKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT
    orf47ng-1    250     260     270     280     290     300
    VQVYRWYKYPVLKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT

50  orf47-1.pep  310     320     330     340     350     360
    LGMMARTALGHTGNPIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
    orf47ng-1    310     320     330     340     350     360
    LGMMARTALGHTGNSIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA

55  orf47-1.pep  370     380
    LALLVYAWKYIPWLIRPRSDGRPGX
    orf47ng-1    370     380
    LALLVYAWKYIPWLIRPRSDGRPGX

```

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

```

65  gnl|PID|e246540 (Z73914) ORF396 protein [Pseudomonas stutzeri] Length = 396
    Score = 155 bits (389), Expect = 5e-37

```

-408-

Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

```

Query: 7  FVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFY-----WHAHEMIWGYAGLV 59
          P+W +AFRPF+  +LY L++ LW  +TG  GF  WH HEM++G+A  +
5  Sbjct: 14 PIWRLAFRPFFLAGSLYALLAIPLVAAWTGLWP--GFQPTGCWLAWHRHEMLFGFAMAI 71

Query: 60 VIAFLLTAVATWTGQPPTRGGVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119
          V  FLLTAV TWTGQ  G  LVGL A WLAAR+  ++ G AA  L  LF
10 Sbjct: 72 VAGFLLTAVQTWTGQTAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPLDLFLVALVW 130

Query: 120 CMALPVIRSONRRNYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXMVSGFIGL 179
          MA  +  + +RNY V  + ++ G  +V+  + L
15 Sbjct: 131 MMAQMLWAVRQKRNYPIVVVLSLMLGADVLIILTGLLQGNDAQRQGVLAGLWLVAALMAL 190

Query: 180 IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV---MPWLSAAFAFA 234
          IG R+I FFT + L  P W+ A L  + A+L A GV  P L  F A
20 Sbjct: 191 IGGRVIPFFTQRGLGKVDVAVKPWWLDVALLVGTGVIALLHAFGVAMRPQPLLGLLFV-A 249

Query: 235 AGVIFTVQVYRWYKPVLPKEPMLWILFAGYLTGLGLIAGVASYF-KPAFXXXXXXXXXXX 293
          GV  +++ RW+ K + K  +LW L  L+  +  +  +F  A
25 Sbjct: 250 IGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLVVAAFGALWLFGLLAQSSPSLHALSV 309

Query: 294 XXXXXXXXXXXMMARTALGHTGNSIYPPPKAVPVAFWLXXXXXXXXXXXXXFSSTAYTHSIR 353
          M+AR LGHTG + P  + AF L  F S  +
30 Sbjct: 310 GSMSGLILAMIARVTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVAWPVGGLW 365

Query: 354 TSSVLFALALLVYAWKYIPWLIRPRSDGRPG 384
          ++V + LA  +Y W+Y P L+  R DG PG
30 Sbjct: 366 LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

```

35      1  ..ATGCCGTCTG AAGGTTTCAGA CGGCmTCGGT GyCGGGGAay CAGAAGyGGT
      51  AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG
     101  CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT
     151  GATGTCGAAa CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC
     201  TGTkGCTTTC GTGATAGGsA GGTtTGyTGG kmksAsyITG TAyrATwkkG
40     251  CCTssCwsTG kAGmGCCkTk CkyTGGTkkA swGrwArTAG TCGTGGTtTy
     301  TkTtyyCACC GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG
     351  CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG
     401  GACAAGCCGA GAGAAGAAAC GGCGTGGAAG CTGCCGTTTC CCTGATGTTT
     451  TGTTTGGGTT TCTTTGTAGT TGTTGTTTAT CTCTTCAGTA ACTTTTTTtag
45     501  TAGAAGAATT ACTTTCTTTC CATTTTCTGT AACTGGCATA ATCTGCCGCT
     551  ATTCTCCAGC CGCCGAAATC ..

```

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

```

      1  ..MPSEGSdGXG XGEXEXVAHA QXDFVGFEG VfqASpVVVT VSGVXXQLGX
     51  DVETDTGDDT KTXAADXVAF VIGRFxGXXL YXXAXXXAX XWXXXSRGF
50    101  XXHRMNLmFN VSGDARADI GFEFIVEFEI VNGGQAERN GVEAAVSLMF
     151  CLGFFVVVVY LFSNFFSRRI TFFPFsvTGI ICRYSpAAEI ..

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from  
55 *N.gonorrhoeae*:

	orf67.pep		MPSEGS	DXGXGEXE	XVAHAQ	XDFVGF	FEAG	30
	orf67ng	TNFEI	AVLSG	MTVRV	FYCAR	PAPVNG	GR	146
		90	100	110	120	130	140	
5	orf67.pep	VFQAS	PVVTV	SGVXX	QLGX	DVETD	TGDDT	90
	orf67ng	VFQAS	PVVAV	AGVQ	QAGRD	VYAHAR	HRAEAQ	206
10	orf67.pep	XWXXX	SRGFX	XHRM	NLMFN	SVSGD	ARADIG	150
		:	:	:	:	:	:	
	orf67ng	TRVGG	KSTCY	FFSR	IDAVS	DVSVG	DARTD	266
15	orf67.pep	CLGFF	VV-----	VVYLF	SNFFS	RRITFF	-PFSVT	190
			:	:	:	:	:	
	orf67ng	RLLVF	YVKLV	AAKS	FIILS	FQLFY	VHGIF	326

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising amino acid sequence <SEQ ID 718>:

	1	MPSET	VGSIV	NVGVD	ESVGF	SPPF	PSIQHF	YRFH	RIHRIR	LFRPP	GPQMQL
20	51	NRHSH	GSGNL	GRGV	WATVLS	DKFP	CGQVRI	PACAG	MTNFE	IAVL	SGMTVR
	101	VFYCAR	PAPV	NGGRL	KMPSE	GSDGI	GIGES	EAVA	HAQORF	VGFE	AGVFQA
	151	SPVVAV	AGV	QQAG	RDVYA	HARH	RAEAQA	AAVA	FLIGV	FLRMS	VRINR
	201	NCCVSI	TRVG	GKST	CYFFSR	IDAVS	DVSVG	DARTD	IGFEF	VVEFE	IVNGG
	251	QAERRN	GVEC	AVFLM	FRLLV	FYVKL	VAAKS	FIILS	FQLFY	VHGIF	FIVVPF
25	301	PVTGI	IRGDA	PAAEV	VADRH	PGVD	GMRTDV	SEIIA	YRAYF	VFAWS	GWFR
	351	IVGNA	FGG	VG			*				

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 30 Example 86

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 719>

	1	ATGTTT	GCTT	TTT	TAGAAGC	CTTTT	TTGTC	GAATA	CGGTT	ATGCG	GGCTGT
	51	TTTTTT	TGTA	TTGGT	CATCT	GCGTT	TCGG	CGTG	CCGATT	CCCGA	GAGGATT
35	101	TGACCT	TGGT	AACAGG	CGGC	GTGAT	TCGG	GTATG	GGTTA	TACCA	ATCCG
	151	CATATT	TATGT	TTGCAG	TCGG	TATGCT	CGGC	GTATT	GGTCG	GGGAC	GGCAT
	201	CATGTT	CGCC	GCCGG	ACGAA	TTTGGG	GGCA	GAAr	ArTCCTA	rGGTT	CArAC
	251	CTATTG	CGsG	CATCAT	GACG	CCGr	AACGTT	ATGAG	CAGGT	TCAGG	AAAAA
	301	TTCGAC	AAAT	ACGGT	AACTG	GGTCT	TATTT	GTCG	CCCGTT	TCCTG	CCCCG
40	351	TTTGAG	AACG	GCCGT	TATTTG	TTACAG	CCCG	TATCA	GCCGC	AAGGT	TTTCAT
	401	ACTTG	CGTTT	TATCAT	TATG	GATGG	ACTGG	CCGCA			

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

	1	MFAF	LEAFFV	EYGYA	AVFFV	LVICG	FVPI	PEDLT	LV	TGG	VISGM	GYTNP
	51	HIMFA	VGM	LV	VDG	IMFA	AGRIW	GQXXL	XFXPI	AXIMT	PXRYE	QVQEK
	101	FDKYG	NWVLF	VAREL	PGLRT	AVFVT	AGISR	KVSYL	RFIIM	DGLAA		

45 Further work revealed the complete nucleotide sequence <SEQ ID 721>:

	1	ATGTTT	GCTT	TTT	TAGAAGC	CTTTT	TTGTC	GAATA	CGGTT	ATGCG	GGCTGT
	51	TTTTTT	TGTA	TTGGT	CATCT	GCGTT	TCGG	CGTG	CCGATT	CCCGA	GAGGATT
	101	TGACCT	TGGT	AACAGG	CGGC	GTGAT	TCGG	GTATG	GGTTA	TACCA	ATCCG
50	151	CATATT	TATGT	TTGCAG	TCGG	TATGCT	CGGC	GTATT	GGTCG	GGGAC	GGCAT
	201	CATGTT	CGCC	GCCGG	ACGAA	TTTGGG	GGCA	GAAA	ATCCTA	AGGTT	CAAAC
	251	CTATTG	CGCG	CATCAT	GACG	CCGAA	ACGTT	ATGAG	CAGGT	TCAGG	AAAAA
	301	TTCGAC	AAAT	ACGGT	AACTG	GGTCT	TATTT	GTCG	CCCGTT	TCCTG	CCCCG
	351	TTTGAG	AACG	GCCGT	TATTTG	TTACAG	CCCG	TATCA	GCCGC	AAGGT	TTTCAT
	401	ACTTG	CGTTT	TATCAT	TATG	GATGG	ACTGG	CCGCA	CTGAT	TTCCG	TCCCT
55	451	ATTTG	GATTT	ATCTG	GGCGA	ATACG	GTGCG	CACA	ACATCG	ATTGG	CTGAT

```

501  GGCAGAAATG CACAGCCTGC AATCGGGTAT TTTTGTATC TTGGGTATAG
551  GTGCGACCGT TGTGCTTGG ATTTGGTGGA AAAACGCCA ACGTATCCAG
601  TTTTACCGCA GCAAATTGAA AGAAAAGCGG GCGCAACGCA AAGCCGCCAA
651  GGCAGCCAAA AAAGCCGCGC AAAGCAAACA ATAA

```

5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

```

1  MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
51  HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL RFKPIARIMT PKRYEQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAALISVP
151 IWIYLGEYGA HNIDWLMAMK HSLQSGIFVI LGIGATVVAW IWWKKRQRIQ
10  201 FYRSKLKEKR AQRKAATAK KAAQSKQ*

```

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H.influenzae* (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

```

15  Orf78: 4  FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGV 61
      FL  FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+      N H+M  V M+GV
      DedA: 20 FLIGFFTEYGYWAVLFLVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGV 79

      Orf78: 62 LVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121
      L GD M+  GRI+G  L F PI  I+T  R  V+EKF +YGN VLFVARFLPGLR
      DedA: 80 LAGDSMYWLGRIGTKILRFRPIRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139

      Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145
      +++ +GI+R+VSY+RF+++D  AA
      DedA: 140 IYMVSGITRRVSVRVFLIDFCAA 163
25

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N. meningitidis*:

```

30  orf78.pep  MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
      orf78a   MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
      10      20      30      40      50      60

35  orf78.pep  VLVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRT
      orf78a   VLVGDGIMFAAGRIWGQKILFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
      70      80      90      100     110     120

40  orf78.pep  AVFVTAGISRKVSYLRFIIMDGLAA
      orf78a   AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLSQSGIFIA
      130     140     150     160     170     180
45

```

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

```

1  ATGTTTGCCC TTTTGAAGC CTTTTTGTG GAATACGGCT ATGCGGCCGT
51  GTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
101 TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCC GGGACGGCAT
201 CATGTTCCGCC GCCGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC
251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCACAGGT TCAGGAAAAA
301 TTCGACAAAT ACGGCAACTG GGTGTTATTT GTCGCTCGTT TCCTGCCCCG
351 TTTGCGGACT GCCGTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT
55  401 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCCGTGCCC
      451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGCGCTGAT

```

501 GGCGAAAATG CACAGCCTGC AATCCGGCAT CTTCATCGCA TTGGGCGTGC  
 551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG  
 601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA  
 651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

5 This encodes a protein having amino acid sequence <SEQ ID 724>:

1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP  
 51 HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK  
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFILM DGLAALISVP  
 151 VWIYLGEYGA HNIDWLMAMK HSLQSGIFIA LGVLAALAW FWRKRRHYQ  
 201 LYRAQLSEKR AKRKAKEAAK KAAQKQQ\*

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

		10	20	30	40	50	60
orf78a.pep		MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNP	HIMFAVGMLG				
15	orf78-1	MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNP	HIMFAVGMLG				
		10	20	30	40	50	60
		70	80	90	100	110	120
20	orf78a.pep	VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT					
	orf78-1	VLVGDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT					
		70	80	90	100	110	120
		130	140	150	160	170	180
25	orf78a.pep	AVFVTAGISRKVSYLRFILMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLSQSGIFIA					
	orf78-1	AVFVTAGISRKVSYLRFIIMDGLAALISVPIWIYLGEYGAHNIDWLMAMKHSLSQSGIFVI					
		130	140	150	160	170	180
30		190	200	210	220		
	orf78a.pep	LGVLAAALAWFWRKRRHYQLYRAQLSEKRAKRAKAAKAAQKQQX					
	orf78-1	LGIGATVVAWIWKKRQRIQFYRSKLKEKRAKRAKAAKAAQSKQX					
35		190	200	210	220		

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

40	orf78.pep	XXLXFPIAXIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF	137
	orf78ng	YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF	32
	orf78.pep	IIMDGLAA	145
45	orf78ng	LIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLSQSGIFIALGVLAALAWFWRKRR	92

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

1 ..YPVLFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPVWIYL  
 51 GEYGAHNIDW LMAKMHSLSQ GIFIALGVLA AALAWFWRK RRHYQLYRAQ  
 101 LSEKRAKRKA EKAAKKAAQK QQ\*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

1 atgtttgccc tttTggaagc CTTTTTTGTC GAAtacggCt atgcGGCCGT  
 51 GTTTTTTCGT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAAGATT  
 101 TGACCTTGGT AACGGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG  
 151 CATATTATGT TTGCGGTCGG TATGCTCGGC GTGTTGGCGG GCGACGGCGT  
 201 GATGTTTGCC GCCGGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC  
 251 CGATTGCCCG CATCATGACG CCGAAACGTT ACGCGCAGGT TCAGGAAAAA  
 301 TTCGACAAAT ACGGCAACTG GGTTCTGTTT GTCGCCCGTT TCCTGCCGGG

-412-

5  
 351 TTTGCGGACT GCCGTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT  
 401 ATCTGCGCTT TCTGATTATG GACGGGCTGG CCGCGCTGAT TTCCGTGCCC  
 451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT  
 501 GGCGAAAATG CACAGCCTGC AATCGGGCAT CTTATCGCA TTGGGCGTGC  
 551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG  
 601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA  
 651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

10  
 1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTlVTGG VISGMGYTNP  
 51 HIMFAVGMLG VLAGDGVMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK  
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFILM DGLAALISVP  
 151 VWIYLGEYGA HNIDWLMAKM HSLQSGIFIA LGVLAAALAW FWRKRHHYQ  
 201 LYRAQLSEKR AKRKA EKA KAAQKQ\*

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

15  
 orf78-1.pep 10 20 30 40 50 60  
 MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTlVTGGVISGMGYTNPHIMFAVGMLG  
 orf78ng-1 MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTlVTGGVISGMGYTNPHIMFAVGMLG  
 20  
 orf78-1.pep 70 80 90 100 110 120  
 VLVDGIMFAAGRIWGQKILREKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT  
 orf78ng-1 VLAGDGVMFAAGRIWGQKILREKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT  
 25  
 orf78-1.pep 70 80 90 100 110 120  
 VLVDGIMFAAGRIWGQKILREKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT  
 orf78ng-1 VLAGDGVMFAAGRIWGQKILREKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT  
 30  
 orf78-1.pep 130 140 150 160 170 180  
 AVFVTAGISRKVSYLRFILMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFVI  
 orf78ng-1 AVFVTAGISRKVSYLRFILMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIA  
 35  
 orf78-1.pep 190 200 210 220  
 LGIGATVVAWIWKKRQRIQFYRSKLKEKRAQRKAAKAAQSKQX  
 orf78ng-1 LGVLAAALAWFWWRKRHHYQLYRAQLSEKRAKRAEKAAKAAQKQX  
 190 200 210 220

Furthermore, orf78ng-1 shows homology to the dedA protein from *H. influenzae*:

40  
 sp|P45280|YG29 HAEIN HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir|D64133 dedA  
 protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)  
 >gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212  
 Score = 223 bits (563), Expect = 7e-58  
 Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)  
 45  
 Query: 5 LEAFFVEYGYAAVFFVLVICGFGVPIPEDLTlVTGGVISGM--GYTNPHIMFAVGMLGVL 62  
 L FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GVL  
 Sbjct: 21 LIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGVL 80  
 50  
 Query: 63 AGDGVMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRTAV 122  
 AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLPGLR +  
 Sbjct: 81 AGDSCMYWLGRIYGTILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 140  
 55  
 Query: 123 FVTAGISRKVSYLRFILMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIALG 182  
 ++ +GI+R+VS+Y+RF+++D AA+ISVP+WIYLGE GA N+DWL ++ Q I+I +G  
 Sbjct: 141 YMVSGITRRVS+YR+VFLIDFCAAII+SVPIWIYLGE+GAKNLDWLHTQIQKGQIVYIFIG 200  
 Query: 183 VL 184  
 L  
 Sbjct: 201 YL 202  
 60



Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 87

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

```

10      1  ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
      51  TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
     101  AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
     151  AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
     201  AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
     251  AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
     301  AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
     351  TAAAATCCCC GTTACCCTGA AATTAAAAAA CGCCAAAGCG CAAACCGTCC
     401  AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA C...
```

15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

```

      1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA
     51  KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
    101  SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNH...
```

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

```

20      1  ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
      51  TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
     101  AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
     151  AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA
     201  AGTGCATACC CACATCAACG ACAACGGCGT GATGCGGATG CGCGAAGTCG
     251  AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
     301  AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
     351  TAAAATCCCC GTTACCCTGA AATTAAAAAA CGCCAAAGCG CAAACCGTCC
     401  AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA CGGTCATCAC
     451  CACGCGGAAG CGCATCAGCA CTAA
```

30 This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

```

      1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA
     51  KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
    101  SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNHGHH
    151  HGEAHQH*
```

35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N.meningitidis*:

```

40      10      20      30      40      50      60
    orf79.pep  MKKLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKIGGAFMKIHNDEAKQDFLLGGSS
    orf79a     MKXLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGMKMGGAFFMKIHNDEAKQDFLLGGSS
      10      20      30      40      50      60
45      70      80      90     100     110     120
    orf79.pep  PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
    orf79a     PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXXKQLKXGDKIP
      70      80      90     100     110     120
50
```

-414-

```

              130      140
orf79.pep    VTLKFKNAKAQTVQLEVKIAFPMPAMNH
              |||||
5  orf79a     VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQH
              130      140      150

```

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

```

1  ATGAAANAAC TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAATCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATGGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCTGTTGCCG ACCGCGTCGA
201 AGTGCATACC CATATCAATG ATAACGGTGT GATGCGGATG CGCGAAGTCG
251 AAGCGGCGGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCGGCG
301 AGCTATCATG TCATGTTTAT GGGTNTGAAA AAACAATTAA AAGANGGCGA
15 351 CAAGATTCCT GTTACCCTGA AATTTAAAAA CGCCAAAGCA CAAACCGTCC
401 AACTGGAAGT CAAAACCGCG CCGATGTCGG CAATGGACCA CGGTCATCAC
451 CACGGCGAAG CGCATCAGCA CTAA

```

This encodes a protein having amino acid sequence <SEQ ID 734>:

```

1  MKXLLAAVMM AGLAGAVSAA GIHVEDGWAR TVEGMKMGGA AFMKIHND
20 51 KQDFLLGGSS PVADRVVHT HINDNGVMMR REVEGGVPLE AKSVTELKPG
101 SYHVMFMGXK KQLKXGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMDHGHH
151 HGEAHQH*

```

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

```

              10      20      30      40      50      60
25 orf79a.pep MKXLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGMKMGGA FMKIHND
    || |||||
orf79-1       MKXLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKIGGA FMKIHND
              10      20      30      40      50      60

30 orf79a.pep PVADRVEVHTHINDNGVMMRREVEGGVPLEAKSVTELKPGSYHVMFMGXK KQLKXGDKIP
    |||||
orf79-1       PVADRVEVHTHINDNGVMMRREVEGGVPLEAKSVTELKPGSYHVMFMGLK KQLKEGDKIP
              70      80      90      100     110     120

35 orf79a.pep VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQH
    |||||
orf79-1       VTLKFKNAKAQTVQLEVKIAFPMPAMNHGHHHGEAHQH
              130     140     150
40 orf79a.pep
    |||||
orf79-1       VTLKFKNAKAQTVQLEVKIAFPMPAMNHGHHHGEAHQH
              130     140     150

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

```

45 orf79.pep FMKIHND EAKQDFLLGGSSPVADRVEVHTHINDNGVMMRREVEGGVPLEAKSVTELKPGS 101
    |||||
orf79ng      INDNGVMMRREVKGGVPLEAKSVTELKPGS 30

50 orf79.pep YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKIAFPMPAMNH 147
    |||||
orf79ng      YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH 86

```

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

```

1  ..INDNGVMMR EVKGGVPLEA KSVTELKPGS YHVMFMGLKK QLKEGDKIPV
55 51 TLKFKNAKAQ TVQLEVKTAP MSAMNHGHHH GEAHQH*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

-415-

```

      1 ATGAAAAAAT TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
      51 TTccgcccGc GGagTccAtG TCGAggACGG CTGGGCGCGc accaCTGtcg
     101 aaggtATgaa aatggGCGGC GCgttCATga aaATCCACAA CGACGaaGcc
     151 atacaaGACt ttgtgcTCgg CGGaagcatg cccgttgccg accgcGTCTGA
     201 AGTGCAAtaca cacATCAACG ACAACGGCGT GATGCGTATG CGCGAAGTCA
     251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGCG
     301 AGCTATCACG TGATGTTTAT GGGTTTGAAA AAACAACTGA AAGAGGGCGA
     351 CAAGATTCCC GTTACCCTGA AATTAAAAA CGCCAAAGCG CAAACCGTCC
     401 AACTGGAAGT CAAAACCGCG CCGATGTCGG CAATGAACCA CGGTCATCAC
     451 CACGGCGAAG CGCATCAGCA CTAA

```

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

```

      1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKMG GAFMKIHND EAKQDFVLG GSS
     51 IQDFVLGGSM PVADRVEVHT HINDNGVMR REVKGVPLE AKSVTELKPG
    101 SYHVMFMGLK KQLKEGDKIP VTLKFKNKA QTVQLEVKTA PMSAMNHGHH
    151 HGEAHQH*

```

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

```

      10      20      30      40      50      60
or79-1.pep MKKLLAAVMMAGLAGAVSAAAGVHVEDGWARTTVEGMKIGGAFMKIHND EAKQDFVLGGSS
      10      20      30      40      50      60
or79ng-1   MKKLLAAVMMAGLAGAVSAAAGVHVEDGWARTTVEGMKMGGA FMKIHND EAIQDFVLGGSM
      10      20      30      40      50      60
      70      80      90     100     110     120
or79-1.pep PVADRVEVHTHINDNGVMRMREV EGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
      70      80      90     100     110     120
or79ng-1   PVADRVEVHTHINDNGVMRMREV KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
      70      80      90     100     110     120
      130     140     150
or79-1.pep VTLKFKNKAQTVQLEVKIAPMPAMNHGHHHGEAHQH X
      130     140     150
or79ng-1   VTLKFKNKAQTVQLEVKTAPMSAMNHGHHHGEAHQH X
      130     140     150

```

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

```

35 gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
   Score = 63.6 bits (152), Expect = 6e-10
   Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)

40 Query: 24 VEDGWARTTVEGMKMGGA FMKIHND EAIQDFVLGGSM PVADRVEVHTHINDNGVMRMREV 83
      V+   W       G       M I N+   D+++G   +A RVE+H   + +N V +M
      Sbjct: 27 VKHPWVMEPPPGPNTTMMGM IIVNEGDEPDYLGAKTDIAQRVELHKTVIENDVAKMVPQ 86

      Query: 84 KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFKNKAQTVQLEV 137
      + + + K   E K   YHVM +GLKK++KEGDK+ V L F+ +   TV+   V
45 Sbjct: 87 ER-IEIPPKGVFEFKHHGYHVM IIGLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

**Example 88**

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 739>:

```

5      1  ATGACGGTAA  CTGCGGCCGA  AGGCGGCAAA  GCTGCCAAGG  CGTTAAAAAA
      51  ATATCTGATT  ACGGGCATTT  TGGTCTGGCT  GCCGATTGCG  GTAACGGTTT
     101  GGGTGGTTTC  CTATATCGTT  TCCGCGTCCG  ATCAGCTCGT  CAACCTGCTG
     151  CCGAAGCAAT  GCGGCGCCGA  ATATGTTTTG  GGGTTTAATA  TCCCGGGGCT
     201  GGGCGTTATC  GTTGCCATTG  CCGTATTGTT  TGTAACCGGA  TTGTTTGCCG
     251  CCAACGTATT  GGGTCGGCAG  ATCCTCGCCG  CGTGGGACAG  CCTGTTGGGG
    10  301  CGGATTCGGG  TTGTGAAATC  CATCTATTCG  AGTGTGAAAA  AAGTATCCGA
     351  ATacgTGCTG  TCCGACAGCA  GCCGTTCGTT  TAAAACGCCG  GTACTCGTGC
     401  CGTTTCCCA  GCCCGGTATT  TGGACGATyG  CTTTCGTGTC  AGGGCAGGTG
     451  TCGAATGCGG  TTAAGGCCGC  ATTGCCGAAs  GACGGCGATT  ATCTTCCGT
     501  GTATGTTCCG  ACCACGCCGA  ATCCGACCGG  CGGTTACTAT  ATTATGGTAA
    15  551  AGAAAAGCGA  TGTGCGCGAA  CTCGATATGA  GCGTGGACGA  AsCATTGAAA
     601  TATGTGATTT  CGCTGGGTAT  GGTCAATCCCT  GACGACCTGC  CCGTCAAAAC
     651  ATTGGCAsGA  CCTATGCCGT  CTGAAAAGGC  GGATTTGCCC  GAACAACAAT
     701  AA

```

This corresponds to the amino acid sequence <SEQ ID 740; ORF98>:

```

20      1  MTVTAAEGGK  AAKALKKYLI  TGILVWLPIA  VTVWVVSIV  SASDQLVNLL
      51  PKQWRPQYVL  GFNIPGLGVI  VAIAVLFVTG  LFAANVLGRQ  ILAAWDSLLG
     101  RIPVVKSIYS  SVKKVSEYVL  SDSSRSFKTP  VLVFPFPQGI  WTIAFVSGQV
     151  SNAVKAALPX  DGDYLSVYVP  TTPNPTGGYY  IMVKKSDVRE  LDMSVDEXLK
     201  YVISLGMVIP  DDLPVKTLAX  PMPSEKADLP  EQQ*

```

25 Further work revealed the complete nucleotide sequence <SEQ ID 741>:

```

30      1  ATGACGGAAC  nTGC GGCCGA  AGGCGGCAAA  GCTGCCAArG  CGTTAAAAAA
      51  ATATCTGATT  ACGGGCATTT  TGGTCTGGCT  GCCGATTGCG  GTAACGGTTT
     101  GGGTGGTTTC  CTATATCGTT  TCCGCGTCCG  ATCAGCTCGT  CAACCTGCTG
     151  CCGAAGCAAT  GCGGCGCCGA  ATATGTTTTG  GGGTTTAATA  TCCCGGGGCT
     201  GGGCGTTATC  GTTGCCATTG  CCGTATTGTT  TGTAACCGGA  TTGTTTGCCG
     251  CCAACGTATT  GGGTCGGCAG  ATCCTCGCCG  CGTGGGACAG  CCTGTTGGGG
     301  CGGATTCGGG  TTGTGAAATC  CATCTATTCG  AGTGTGAAAA  AAGTATCCGA
     351  ATCGCTGCTG  TCCGACAGCA  GCCGTTCGTT  TAAAACGCCG  GTACTCGTGC
     401  CGTTTCCCA  GCCCGGTATT  TGGACGATTG  CTTTCGTGTC  AGGGCAGGTG
     451  TCGAATGCGG  TTAAGGCCGC  ATTGCCGAAG  GACGGCGATT  ATCTTCCGT
     501  GTATGTTCCG  ACCACGCCGA  ATCCGACCGG  CGGTTACTAT  ATTATGGTAA
     551  AGAAAAGCGA  TGTGCGCGAA  CTCGATATGA  GCGTGGACGA  AGCATTGAAA
     601  TATGTGATTT  CGCTGGGTAT  GGTCAATCCCT  GACGACCTGC  CCGTCAAAAC
     651  ATTGGCAGGA  CCTATGCCGT  CTGAAAAGGC  GGATTTGCCC  GAACAACAAT
    40  701  AA

```

This corresponds to the amino acid sequence <SEQ ID 742; ORF98-1>:

```

45      1  MTEXAAEGGK  AAKALKKYLI  TGILVWLPIA  VTVWVVSIV  SASDQLVNLL
      51  PKQWRPQYVL  GFNIPGLGVI  VAIAVLFVTG  LFAANVLGRQ  ILAAWDSLLG
     101  RIPVVKSIYS  SVKKVSESL  SDSSRSFKTP  VLVFPFPQGI  WTIAFVSGQV
     151  SNAVKAALPK  DGDYLSVYVP  TTPNPTGGYY  IMVKKSDVRE  LDMSVDEALK
     201  YVISLGMVIP  DDLPVKTLAG  PMPSEKADLP  EQQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF98 shows 96.1% identity over a 233aa overlap with an ORF (ORF98a) from strain A of *N.*

50 *meningitidis*:

```

55      orf98.pep      10      20      30      40      50      60
      orf98a      MTVTAAEGGKA AAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLL PKQWRPQYVL
      || |||||
      MTPEAAEGGKA AAKALKKYLI TGILVWLPIA VTVWVVSIV SASDQLVNLL PKQWRPQYVL
      10      20      30      40      50      60

```

-417-

		70	80	90	100	110	120
	orf98.pep	GFNIPGLGVIVAI	AVLFVTGLFAAN	VLGRQILAAWDS	LLGRIPVVKSIY	SSVKKVSEYVL	
5	orf98a	GFNIPGLGVIVAI	AVLFVTGLFAAN	VLGRQILAAWDS	LLGRIPVVKSIY	SSVKKVXSLL	
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf98.pep	SDSSRSFKTPVL	VFPFQPGIWTIA	FSVGQVSNVKA	ALPKDGDYLSV	YVPTTPNPTGG	YY
	orf98a	SDSSRSFKTPVL	VFPFQSGIWTIA	FSVGQVSNVKA	ALPKDGDYLSV	YVPTTPNPTGG	YY
		130	140	150	160	170	180
		190	200	210	220	230	
15	orf98.pep	IMVKKSDVRELD	MSVDEALKYVIS	LGMVIPDDL	VPVKTLAGPMP	SEKADLPEQQX	
	orf98a	IMVKKSDVRELD	MSVDEALKYVIS	LGMVIPDDL	VPVKTLAGPMP	SEKADLPEQQX	
		190	200	210	220	230	

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

20	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCCAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTGT	GGGTTTAATA	TCCCGGGGCT
	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTGCGG
25	251	CAAACGTATT	GGGCCGGCAG	ATTCTTGCCG	CGTGGGACAG	CTTGTGCGG
	301	CGGATTCGGG	TTGTGAAGTC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCCGTTTCGT	TAAAACACCA	GTACTCGTGC
	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCGTGTC	CGGTCAGGTG
	451	TCAATGCGG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTCCGT
30	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGTTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTTGCCC	GAACAACAAT
	701	AA				

35 This encodes a protein having amino acid sequence <SEQ ID 744>:

	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAVLFVVTG	LFAANVLGRQ	ILAAWDSLLG
	101	RIPVVKSIYS	SVKKVXSLL	SDSSRSFKTP	VLVFPFQSGI	WTIAFVSGQV
	151	SNVKAALPK	DGDYLSVYVP	TFNPTGGYY	IMVKKSDVRE	LDMSVDEALK
40	201	YVISLGMVIP	DDLVPVKTLAG	PMPSEKADLP	EQQ*	

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

		10	20	30	40	50	60
	orf98a.pep	MTEPAAEGGKA	AKALKKYLTG	ILVWLPIAVT	VWVSVIVSAS	DQLVNLLPKQ	WRPQYVL
45	orf98-1	MTEXAAEGGKA	AKALKKYLTG	ILVWLPIAVT	VWVSVIVSAS	DQLVNLLPKQ	WRPQYVL
		10	20	30	40	50	60
		70	80	90	100	110	120
50	orf98a.pep	GFNIPGLGVIV	AI	AVLFVTGLFA	ANVLGRQILAA	WDSLLGRIPV	VKSIYSSVKKV
	orf98-1	GFNIPGLGVIV	AI	AVLFVTGLFA	ANVLGRQILAA	WDSLLGRIPV	VKSIYSSVKKV
		70	80	90	100	110	120
		130	140	150	160	170	180
55	orf98a.pep	SDSSRSFKTPV	LVFPFQSGI	WTIAFVSGQV	SNVKAALPKD	GDYLSVYVPT	TPNPTGGYY
	orf98-1	SDSSRSFKTPV	LVFPFQSGI	WTIAFVSGQV	SNVKAALPKD	GDYLSVYVPT	TPNPTGGYY
		130	140	150	160	170	180
		190	200	210	220	230	
60	orf98a.pep	IMVKKSDVREL	DMSVDEALKY	VISLGMVIPD	DLVPVKTLAG	PMPSEKADLP	EQQX
	orf98-1	IMVKKSDVREL	DMSVDEALKY	VISLGMVIPD	DLVPVKTLAG	PMPSEKADLP	EQQX
		190	200	210	220	230	

65

Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

5	orf98.pep	MTVTAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
	orf98ng	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
10	orf98.pep	GFNIPGLGVIVAI AVL FVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL	120
	orf98ng	GFNIPGLGVIVAI AVL FVTGLFAANVLGRQILAAWDSLLXRI PVVKSIYSSVKKVSESL	120
15	orf98.pep	SDSSRSFKTPVLVPFPQPGIWTIAFVSGQVSNVKAALPXDGDYLSVYVPTTNPNTGGYY	180
	orf98ng	SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNVKAALPDGDYLSVYVPTTNPNTGGYY	180
20	orf98.pep	IMVKKSDVRELDMSVDEXLKYVISLGMVIPDDL PVKTLAXPMPSEKADLPEQQ	233
	orf98ng	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDL PVKTLAGPMPPEKAELPEQQ	233

- 20 The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

25	1	MTEPAAEGGK	AAKALKKKYL	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLX
	101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPFPQSGI	WTIAFVSGQV
	151	SNVKAALPQ	DGDYLSVYVP	TPNPNTGGYY	IMVKKSDVRE	LDMSVDEALK
	201	YVISLGMVIP	DDL PVKTLAG	PMPPEKAELP	EQQ*	

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

30	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACAGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ACCAGCTTGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCCGGGCT
	201	CGGCGTTATT	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTGCCG
35	251	CAAACGTGTT	GGGCCGCGAG	ATTCTTGCCG	CGTGGGACAG	CCTGTTgggg
	301	cgaATTCCGG	TTGTCAAATC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
	351	ATCGCTGCTG	TCCGACAGCA	GCCGTTCTGT	TAAAACGCCG	GTA CTCTGTC
	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCGTGTC	CGGTCAGGTG
	451	TCCAATGCGG	TTAAGGCCGC	ATTGCCGCGG	GATGGCGATT	ATCTTCCGT
40	501	GTATGTCCCG	ACCACGCCCA	ACCCGACCGG	CGGTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGC	CTGAAAAGGC	GGAGTTGCC	GAACAACAAT
	701	AA				

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

45	1	MTEPAAEGGK	AAKALKKKYL	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLG
	101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPFPQSGI	WTIAFVSGQV
	151	SNVKAALPQ	DGDYLSVYVP	TPNPNTGGYY	IMVKKSDVRE	LDMSVDEALK
	201	YVISLGMVIP	DDL PVKTLAG	PMPPEKAELP	EQQ*	

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

50	orf98-1.pep	MTExAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
	orf98ng-1	MTEPAAEGGKAALKKKYLITGILVWLPIAVTVWVVSIVSASDQLVNLLPKQWRPQYVL	60
55	orf98-1.pep	GFNIPGLGVIVAI AVL FVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL	120
	orf98ng-1	GFNIPGLGVIVAI AVL FVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL	120

-419-

```

    orf98ng-1      GFNIPGLGVIVAVIAVLFTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSVKKVSESL
                    70      80      90      100     110     120
5   orf98-1.pep    SDSSRSFKTPVLVPFPQPGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY
                    130     140     150     160     170     180
    orf98ng-1      SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNVKAALPQDGDYLSVYVPTTPNPTGGYY
                    130     140     150     160     170     180
10  orf98-1.pep    IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPSEKADLPEQQX
                    190     200     210     220     230
    orf98ng-1      IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPKTLAGPMPPEKAELPEQQX
                    190     200     210     220     230

```

- 15 Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 89

- 20 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

```

1   ATgAAAACGG TAGTCTGGAT TGTCGTCCTG TTTGCCGCCG CCGTCGGACT
51  GCGCTGGCT TCGGGCATT TACACGGCGA CGTGTATATC GTACTCGGAC
101 AGACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTTAGG TTCGCTGATT
25 151 GCCGTCGTGG TGTGGTATTT CTTGTTTAAA TTCATTATCG GgGgTACTCA
    201 ATATCCCCGA AAAGATGCAG CGTTTCGGTT CGGCnCGTAA AGGCCkCAAG
    251 ssCGsGCTTG CCTTGAACAA GGCGGGTTTG GCGTATTTTG AAGGGCGTTT
    301 TGAAAAGGCG GAACTAGAAG CCTCACGCGT GTTGGTCAAC AAAGtAGGCC
    351 GaGAGACAAC CGGACTTTGG CATTGATGCT GrGCGCGCAC GCCGCCGAC
    401 AGATGGAAAA CATCGAsTG CGCGACCGTT ATCTTGCGBA AATCGCCAAA
30 451 CTGCCGGAAC AACAGCAGCT TTCCCGTTAT CTTTTGTTGG CGGAATCGGC
    501 GTTGAACCGG CGCGATTACG AAGCGGCGGA AGCCAATCTT CATGCGGCGG
    551 CGAAGATGAA TGCCAACCTT ACGCGCCTCG TCGCTCTGCA .ATTCTGTAC
    601 GCTTTCGACA GGGGCGACGC GTTGCAAGTT CTGGCAAAAA CCGAAAAACT
35 651 TTCCAAGGCG GGCGCGTTGG GCAAAATCGGA AATGGAACGG TATCAAAATT
    701 GGGCATATCC GTCCGCCAGT GGCGGATGCT GCCGATGCCG CCGCTTTGAA
    751 AACCTGCCTG AAGCGGATTC CCGACAGCCT CAAAAACGGG GAATTGAGCG
    801 TATCGGTTGC GGAAGAGTAC GAACGTTTGG GACTGTATGC CGATGCGGTC
    851 AAATGGGTCA AACAGCATT TCCGCAsAAC CGCCGCCCGG AGCTTTTGGA
40 901 AGCCTTTGTC GAAAGCGTGC GCTTTTGGG CGAGCGCGAA CAGCAGAAAG
    951 CCATCGATT TGCCTGATG TGGCTGAAAG AACAGCCCGA TAACGCGCTT
1001 CTGCTGATG ATCTCGGTCG GCTCGCCTTC GGCCGCAAAAC TTTGGGGCAA
1051 GGCAAAAGGC TACCTTGAAG CGAGCATTGC ATTAAGCCG AGTATTTCCG
1101 CCGCTTTGGT TCTAACAAAG GTTTTCGACG AAATCGGAGA ACCGCAGAAG
1151 GCGGAGGCGC AC...

```

- 45 This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

```

1   MKTVVWIVVL FAAAVGLALA SGIYTGdVYI VLGQTMlRIN LHAFVLGSLI
51  AVVVWYFLFK FIIGVLNIPE KMQRFGSARK GXKXXLALNK AGLAYFEGFR
101 EKAELEASRV LVNKVGRDNR TLALMLXAHA AGOMENIXXR DRYLAEIAKL
50 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAAMNANLT RLVLRLXIRYA
    201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT
    251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKVQHYP XNRRPELLEA
    301 FVESVRFLGE REQQKAIDFA DAWLKEQPDN ALLLMYLGRl AFGRKLWGKA
    351 KGYLEASIAL KPSISARLVL TKVFDEIGEP QKAEAH...

```

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

```

55 1   ATGAAAACGG TAGTCTGGAT TGTCGTCCTG TTTGCCGCCG CCGTCGGACT
    51  GCGCTGGCT TCGGGCATT TACACGGCGA CGTGTATATC GTACTCGGAC
    101 AGACCATGCT CAGAATCAAC CTGCACGCCT TTGTGTTAGG TTCGCTGATT
    151 GCCGTCGTGG TGTGGTATTT CTTGTTTAAA TTCATTATCG GCGTACTCAA

```

-420-

201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GCGCGTAAA GGCCGCAAGG  
 251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTT  
 301 GAAAAGGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGGCCGG  
 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGGACAGA  
 401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG  
 451 CCGGAAAAAC AGCAGCTTTC CCGTTATCTT TTGTGGCGG AATCGGCGTT  
 501 GAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA  
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT  
 601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAAGTTTC  
 651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATGGG  
 701 CATACGCGCG CCAGCTGGCG GATGCTGCGG ATGCGCGCGC TTTGAAAACC  
 751 TGCTGAAGC GGATTCCTGA CAGCCTCAA AACGGGAAT TGAGCGTATC  
 801 GGTTCGCGAA AAGTACGAAC GTTGGGACT GTATGCCGAT GCGGTCAAAT  
 851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC  
 901 TTTGTGCGAA GCGTGCCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT  
 951 CGATTTTGCC GATGCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC  
 1001 TGATGTATCT CGGTGCGCTC GCCTACGCC GCAAACTTTG GGGCAAGGCA  
 1051 AAAGGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTTCCGCGCG  
 1101 TTTGTTCTA GCAAAGGTTT TCGACGAAAT CGGAGAACCG CAGAAGGCGG  
 1151 AGGCGCAGCG CAACTTGGTT TTGAAGCCG TCTCCGATGA CGAACGTCAC  
 1201 GCAGCGTTAG AGCAGCATAG CTGA

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

1 MKTVVWIVVL FAAAVGLALA SGIYTGdVYI VLQGTMLRIN LHAFLVLSLI  
 51 AVVVVYFLFK FIIGVNIPE KMQRFGSARK GRKAALALNK AGLAYFEGRF  
 101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAETAKL  
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAKMNNALNT RLVRLQLRYA  
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT  
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYH HNNRRPELLEA  
 301 FVESVRFLGE REQQAIDFA DAWLKEQPDN ALLMYLGR LAYGRKLWGKA  
 351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEAQRLNV LEAVSDDERH  
 401 AALEQHS\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N.*

35 *meningitidis*:

		10	20	30	40	50	60
orf100.pep		MKT	VVW	IVVL	FAAAVGLALAS	SGIY	TGdVYI
orf100a		MKT	VVW	IVVL	FAAAXGLALAS	SGIX	TGdVYI
40		10	20	30	40	50	60
		70	80	90	100	110	120
orf100.pep		FIIG	VLN	IP	EKMQ	RFGS	ARKG
orf100a		FIIG	VLN	XPE	KMQ	RFGS	ARKG
45		70	80	90	100	110	120
		130	140	150	160	170	180
orf100.pep		TLAL	MLX	AHA	AGQ	MEN	IXXR
orf100a		TLAL	MLG	AHA	AGQ	MEN	IELR
50		130	140	150	160	170	180
		190	200	210	220	230	240
orf100.pep		AAAK	MNAN	LTRL	VRLX	IRYA	FDRG
orf100a		AAAK	MNAN	LTRL	VRLX	QLRY	AFDR
55		190	200	210	220	230	240
		250	260	270	280	290	300
orf100.pep		DAAD	AAAL	KTCL	KRIP	DSLK	NGEL
orf100a		DAAD	AAAL	KTCL	KRIP	DSLK	NGEL
60		250	260	270	280	290	300



-421-

		310	320	330	340	350	360
5	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGRILAFGRKLWGKAKGYLEASIAL					
	orf100a	FVESVRFLGERDQQKAIDFADAWLKEQPDNALLXYLGRILAYGRKLWGKAKGYLEASIAL					
		310	320	330	340	350	360
10	orf100.pep	KPSISARLVLTKVFDEIGEPQKAEAH					
	orf100a	KPSISARLVLAKVFDETGEQKAEQARNLVLASVAEENRPSAETHX					
		370	380	390	400		

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

15	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CNNTCGGGCT
	51	GGCATTGGCG	TCGGGCATTN	ACACCGGCGA	CGTGATATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTGCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTCAAA	TTTCATCATCG	GCGTACTCAA
	201	TANCCCCGAA	AAGATGCAGC	GTTTCGGTTC	GGCGCGTAAA	GGCCGCAAGG
	251	CCGCGCTTGC	TTTGAACAAG	GCGGGTTTGG	CGTATTTTGA	AGGGCGTTTT
20	301	GAAGAGGCGG	AACTTGAAGC	CTCGCGCGTA	TTGGGAAACA	AAGAGGCGGG
	351	GGATAACCGG	ACTTTGGCAT	TGATGTTGGG	CGCACATGCC	GCCGGGCAGA
	401	TGGAAAACAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAACTG
	451	CCGAAAAGC	AGCAGCTTTC	CCGTTATCTT	TTGTGCGCG	AATCGGCGTT
	501	GAACCGGCGC	GATTACGAAG	CGGCGGAAGC	CAATCTTCAT	GCGGCGGCGA
25	551	AGATGAATGC	CAACCTTACG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCT
	601	TTTGACAGGG	GCGACGCGTT	GCAGGTTCTG	GCAAAAACCG	AAAAANTTTC
	651	CAAGCGGGC	GCGTNGGGCA	AATCGGAAAT	GGACGGTAT	CAAAATGGGG
	701	CATACCGCCG	CCAGCTGNCG	GATGCTGCCG	ATGCCGCCGC	TTTGAAAACC
	751	TGCCTGAAGC	GGATTCCCGA	CAGCCTCAAA	AACGGGGAAT	TGAGCGTATC
30	801	GGTTGCGGAA	AAGTACGAAC	GTTTGGGACT	GTATGCCGAT	GCGGTCAAAT
	851	GGGTCAAACA	GCATTATCCG	CACAACCGCC	GACCCGAAT	TTTGGAAGCN
	901	TTTGTGCAAA	GCGTGCCTT	TTTGGGCGAA	CGCGATCAGC	AGAAAGCCAT
	951	CGATTTTGCC	GATGCTTGGC	TGAAAGAACA	GCCGATAAT	GCGCTTCTGC
	1001	TGANGTATCT	CGGTGCGCTC	GCCTACGGCC	GCAAACTTTG	GGGCAAGGCA
35	1051	AAAGGCTACC	TTGAAGCGAG	CATTGCATTA	AAGCCGAGTA	TTTCCGCGCG
	1101	TTTGGTTCTG	GCAAAGGTTT	TTGACGAAAC	CGGAGAACCG	CAGAAGGCGG
	1151	AGGCGCAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCCGAGGA	AAACCGNCCT
	1201	TCCGCCGAAA	CCCATTGA			

This encodes a protein having amino acid sequence <SEQ ID 754>:

40	1	MKTVVWIVVL	FAAAXGLALA	SGIXTGDVYI	VLGQTMLRIN	LHAFVLGSLI
	51	AVVVWYFLFK	FIIGVLNXPE	KMQRFSGSARK	GRKAALALNK	AGLAYFEGRF
	101	EKAELEASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLQLRYA
45	201	FDRGDALQVL	AKTEKXSKAG	AXGKSEMERY	QNWAYRRQLX	DAADAALKT
	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWVKQHYP	HNRRPELLEA
	301	FVESVRFLGE	RDQQKAIDFA	DAWLKEQPDN	ALLLXYLGRIL	AYGRKLWGKA
	351	KGYLEASIAL	KPSISARLVL	AKVFDETGE	QKAEQARNLV	LASVAEENRP
	401	SAETH*				

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

50		10	20	30	40	50	60
	orf100a.pep	MKTVVWIVVLFAAAXGLALASGIXTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK					
	orf100-1	MKTVVWIVVLFAAAVGLALASGIYTGVDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK					
55		10	20	30	40	50	60
	orf100a.pep	FIIGVLNXPEKMQRFSGSARKGRKAALALNKAGLAYFEGREFEAELEASRVLGNKEAGDNR					
	orf100-1	FIIGVLNIPEKMQRFSGSARKGRKAALALNKAGLAYFEGREFEAELEASRVLVNKEAGDNR					
60		70	80	90	100	110	120
	orf100a.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
	orf100-1	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
65		130	140	150	160	170	180
	orf100a.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
	orf100-1	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					

-422-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf100a.pep	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKXSKAGAXGKSEMERYQNWAYRRQLX					
	orf100-1	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA					
		190	200	210	220	230	240
10	orf100a.pep	DAADAAALKTKLRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPHNRRPELLEA					
	orf100-1	DAADAAALKTKLRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPHNRRPELLEA					
		250	260	270	280	290	300
15	orf100a.pep	FVESVRFLGERDQQKAIDFADAWLKEQPDNALLXLYLGRLAYGRKLWGKAKGYLEASIAL					
	orf100-1	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL					
		310	320	330	340	350	360
20	orf100a.pep	KPSISARLVLAQVDETGEPQKAEAQRNVLASVAEENRPSA-ETHX					
	orf100-1	KPSISARLVLAQVDEIGEPQKAEAQRNVLAVSDDERHAALQHSX					
25		370	380	390	400		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from

*N.gonorrhoeae*:

30	orf100.pep	MKTVVWIVVLFAAAVGLALASGIYTGVDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK	60
	orf100ng	MKTVVWIVVLFAAAVGLALASGIYTGVDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK	60
35	orf100.pep	FIIGVLNIPEKMQRFSGSARKGXKXXLALNKAGLAYFEGRFEKAELEASRVLVNKVGRDNR	120
	orf100ng	FIIGVLNIPENMRRSGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLGNKEAGDNR	120
40	orf100.pep	TLALMLXAHAAQGMENIXRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
	orf100ng	TLALMLGAHAAQGMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH	180
45	orf100.pep	AAAKMNANLTRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA	240
	orf100ng	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA	240
50	orf100.pep	DAADAAALKTKLRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPXNRRPELLEA	300
	orf100ng	DAADAAALKTKLRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYPHNRRPELLEA	300
55	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL	360
	orf100ng	FVESVRFLGEREQQKAIDFADSWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL	360
	orf100.pep	KPSISARLVLTQVDEIGEPQKAEAH	386
	orf100ng	KPSIPARLVLAQVDETAQSQAQARNVLASVAGENRPSAETR	405

The complete length ORF100ng nucleotide sequence &lt;SEQ ID 755&gt; is:

	1	ATGAAAACGG	TAGTCTGGAT	TGTTGTCCTG	TTTGCCGCCG	CCGTCGGACT
60	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTTAAA	TTCATCATCG	GCGTACTCAA
	201	TATCCCCGAA	AATATGCGGC	GTTCCGGTTC	GGCGCGGAAA	GGCCGCAAGG
	251	CCGCGCTTGC	CTTGAATAAG	GCGGGTTTGG	CGTATTTCTGA	AGGGCGTTTT
	301	GAAAAGGCGG	AACTCGAAGC	CTCTCAGTG	TTGGGCAACA	AAGAGGCCGG
65	351	AGACAACCGG	ACTTTGGCAT	TGATGCTGGG	CGCGCACGCG	GCAGGACAGA
	401	TGGAAAATAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAACTG

-423-

5  
10  
15

```

451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTGGCGG AATCGGCGTT
501 AAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCC
601 TTCGATCGGG GCGATGCGTT GCAGGTCTTG GCAAAAaccG AAAAAGTTTC
651 CAAGCGGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
701 CATACCGCCG CCAGATGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
751 TGCCTGAAGC GGATTCCCGA CAGCCTCAAA AACGGGGAAT TGagcGTATC
801 GGTTCGGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC
901 TTTGTCGAAA GCGTGGCCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
951 CGATTTTGCC GATTCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC
1001 TGATGTATCT CGGCCGGCTC GCCTACGCC GCAAACCTTG GGGTAAGGCA
1051 AAAGGCTACC TTGAAGCGAG TATTGCACTG AAGCCGAGTA TTCCGGCGCG
1101 TTTGCTGTTG GCAAAGGTTT TTGACGAAAC CGCACAGTCG CAAAAGCCG
1151 AAGCACAGCG CAACCTTGGT TTGGCAAGCG TTGCCGGGGA AAACCGCCCT
1201 TCCGCCGAAA CCCGTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 756>:

20  
25

```

1  MKTVVWIVVL FAAAVGLALA SGIYTG DVYI VLQOTMLRIN LHAFVLGSLI
51  AVVVWYFLFK FIIIGVLNIPE NMRRSGSARK GRKAALALNK AGLAYFEGRF
101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
151 PEKQQLSRYL LLAESALNRR DYEAEEANLH AAAMNANLT RLVRQLRYA
201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADAAALKT
251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWVKQHYP HNRPELLEA
301 FVESVRFLGE REQQAIDFA DSWLKEQPDN ALLMYLGR LAYGRKLWGKA
351 KYLEASIAL KPSIPARLV AKVFDETAQS QKAEAQRLV LASVAGENRP
401 SAETR*

```

ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

30  
35  
40  
45  
50  
55  
60  
65

```

      10      20      30      40      50      60
orf100-1.pep  MKTVVWIVVLF FAAAVGLALASGIYTG DVYI VLQOTMLRIN LHAFVLGSLI AVVVWYFLFK
               |||
orf100ng      MKTVVWIVVLF FAAAVGLALASGIYTG DVYI VLQOTMLRIN LHAFVLGSLI AVVVWYFLFK
               |||
      10      20      30      40      50      60

      70      80      90     100     110     120
orf100-1.pep  FIIIGVLNIPE KMQRFGSARK GRKAALALNK AGLAYFEGR FEKAELEASRV LGNKEAGDNR
               |||
orf100ng      FIIIGVLNIPENMRRSGSARK GRKAALALNK AGLAYFEGR FEKAELEASRV LGNKEAGDNR
               |||
      70      80      90     100     110     120

      130     140     150     160     170     180
orf100-1.pep  TLALMLGAHAAGQMENIELR DRYLAEIAKL PEKQQLSRYL LLAESALNRR DYEAEEANLH
               |||
orf100ng      TLALMLGAHAAGQMENIELR DRYLAEIAKL PEKQQLSRYL LLAESALNRR DYEAEEANLH
               |||
      130     140     150     160     170     180

      190     200     210     220     230     240
orf100-1.pep  AAAMNANLTRLVRQLRYAFDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA
               |||
orf100ng      AAAMNANLTRLVRQLRYAFDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA
               |||
      190     200     210     220     230     240

      250     260     270     280     290     300
orf100-1.pep  DAADAAALKTKLRIPDSLK NGELSVSVAE KYERLGLYAD AVKWVKQHYP HNRPELLEA
               |||
orf100ng      DAADAAALKTKLRIPDSLK NGELSVSVAE KYERLGLYAD AVKWVKQHYP HNRPELLEA
               |||
      250     260     270     280     290     300

      310     320     330     340     350     360
orf100-1.pep  FVESVRFLGEREQQAIDFADAWLKEQPDN ALLMYLGR LAYGRKLWGKAKGYLEASIAL
               |||
orf100ng      FVESVRFLGEREQQAIDFADAWLKEQPDN ALLMYLGR LAYGRKLWGKAKGYLEASIAL
               |||
      310     320     330     340     350     360

      370     380     390     400
orf100-1.pep  KPSISARLV LAKVFDEIGE PQKAE AQRLV LA VSDDERHAA LEQHSX
               |||
orf100n       KPSIPARLV LAKVFDETAQS QKAE AQRLV LA VSVAGENRPSAETR*
               |||

```

370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

10      1  ATGATGTTTT  CTTGGTTCAA  GCTGTTTCAC  TTGTTTTTTG  TCATTTCGTG
      51  GTTTCAGGG  CTGTTTACC  TGCCGAGGAT  TTTCGTCAAT  ATGGCGATGA
      101  TTGATGTGCC  GCGCGGCAAT  CCCGAGTATG  TCGCTCTGTC  GGGCATGGCG
      151  GTGCGGCTGT  ACCGTTTAT  GTCGCCGTTG  GGCTTCGGCG  CGGTCGTGTT
      201  CCGCGCGGCG  ATACCGTTT  CCGCCGGCTG  GTGGGGCAGC  GGCTGGGTAC
      251  ACGTCAAAC  GTGTTGGGC  TTGATGCTCT  TGGCTTACCA  GTTGTATTGC
      15  301  GCGGTGCTGC  TGCGCCGTT  TCAGGATTAC  AGCAATGCTT  TTTCACACCG
      351  CTGGTACCG  GTGTTCAAC  AAATCCCGT  GCTGCTGATG  GTTGCCGCGC
      401  TGTATsTGGT  CGTGTTCAA  CCGTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

20      1  MMFSWFKLFH  LFFVISWFAG  LFYLPRI FVN  MAMIDVPRGN  PEYVRLSGMA
      51  VRLYRFMSPL  GFGAVVFGAA  IPFAAGWWGS  GWVHVKLCLG  LMLLAYQLYC
      101  GVLLRRFQDY  SNAFSHRWYR  VFNEIPVLLM  VAALYXVVEK  PF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

25      1  ATGATGTTTT  CTTGGTTCAA  GCTGTTTCAC  TTGTTTTTTG  TCATTTCGTG
      51  GTTTCAGGG  CTGTTTACC  TGCCGAGGAT  TTTCGTCAAT  ATGGCGATGA
      101  TTGATGTGCC  GCGCGGCAAT  CCCGAGTATG  TCGCTCTGTC  GGGCATGGCG
      151  GTGCGGCTGT  ACCGTTTAT  GTCGCCGTTG  GGCTTCGGCG  CGGTCGTGTT
      201  CCGCGCGGCG  ATACCGTTT  CCGCCGGCTG  GTGGGGCAGC  GGCTGGGTAC
      251  ACGTCAAAC  GTGTTGGGC  TTGATGCTCT  TGGCTTACCA  GTTGTATTGC
      301  GCGGTGCTGC  TGCGCCGTT  TCAGGATTAC  AGCAATGCTT  TTTCACACCG
      10  351  CTGGTACCG  GTGTTCAAC  AAATCCCGT  GCTGCTGATG  GTTGCCGCGC
      401  TGTATCTGGT  CGTGTTCAA  CCGTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

35      1  MMFSWFKLFH  LFFVISWFAG  LFYLPRI FVN  MAMIDVPRGN  PEYVRLSGMA
      51  VRLYRFMSPL  GFGAVVFGAA  IPFAAGWWGS  GWVHVKLCLG  LMLLAYQLYC
      101  GVLLRRFQDY  SNAFSHRWYR  VFNEIPVLLM  VAALYLVEFK  PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

40      orf102  3  FSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA VRLYRFMSPLGF 62
      F W K FH+  VISW A LFYLP R+ FV A + V++ +LY F++
      HP1484  8  FLWVKAFHVI AIVISWMAAL FYLPRL FVYHAENAHKKE FVG VVQIQEK--KLYSFIASPA M 65

      orf102  63  GAVVFGAA IPFAAG---WWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDY SNAFSHRWY 119
      G + + + + GW+H KL L ++LLAY YC +R + + R+Y
      45  HP1484  66  GFTLITGILMLLIEPTLFKSGGWLHAKLALVLLLAYH FYCKKCMRELEKDPTRNARFY 125

      orf102  120  RVFNEIPXXXXXXXXXXXXXFKPF 142
      RVFNE P KPF
      HP1484  126  RVFNEAPTILMILIVILVVVKPF 148

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N.*

*meningitidis*:

```

5          10      20      30      40      50      60
   orf102.pep  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMAVRLYRFMSPL
   orf102a     MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMAVRLYRFMSPL
10          10      20      30      40      50      60
   orf102.pep  GFGAVVFGAAIPFAAGWGWGSGWVHV KLC LGLMLLAYQLYCGVLLRRFQDYSNAF SHRWYR
   orf102a     GFGAVVFGAAIPFAAGWGWGSGWVHV KLC LGLMLLAYQLYCGVLLRRFQDYSNAF SHRWYR
15          70      80      90      100     110     120
   orf102.pep  VFNEIPVLLMVAALYLVVFKPFX
   orf102a     VFNEIPVLLMVAALYLVVFKPFX
20          130     140
   orf102.pep  VFNEIPVLLMVAALYLVVFKPFX
   orf102a     VFNEIPVLLMVAALYLVVFKPFX
                130     140

```

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

```

1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCGTG
25 51  GTTTGCAGGG CTGTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
101 TTGATGTGCC GCGCGGCAAT CCCGAGTATG TGCGTCTGTC GGCATGCGC
151 GTGCGGCTGT ACCGTTTTAT GTCGCCGTG GGCTTCGGCG CGGTCGTGTT
201 CCGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGCGAGC GGCTGGGTAC
251 ACGTCAAAC GTGTTTGGGC TTGATGCTCT TGGCTTACCA GTTGTATTGC
301 GCGGTGCTGC TGCGCCGTT TCAGGATTAC AGCAATGCTT TTTCACACCG
351 CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
401 TGTATCTGGT CGTGTTCAAA CCGTTTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 762>:

```

1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
35 51  VRLYRFMSPL GFGAVVFGAA IPFAAGWGS GWVHV KLC LGLMLLAYQLY
101 GVLLRRFQDY SNAF SHRWYR VFNEIPVLLM VAALYLVVFK PF*

```

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

```

          10      20      30      40      50      60
   orf102a.pep  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMAVRLYRFMSPL
40  orf102-1     MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMAVRLYRFMSPL
          10      20      30      40      50      60
   orf102a.pep  GFGAVVFGAAIPFAAGWGWGSGWVHV KLC LGLMLLAYQLYCGVLLRRFQDYSNAF SHRWYR
45  orf102-1     GFGAVVFGAAIPFAAGWGWGSGWVHV KLC LGLMLLAYQLYCGVLLRRFQDYSNAF SHRWYR
          70      80      90      100     110     120
   orf102a.pep  VFNEIPVLLMVAALYLVVFKPFX
   orf102-1     VFNEIPVLLMVAALYLVVFKPFX
50          130     140
   orf102a.pep  VFNEIPVLLMVAALYLVVFKPFX
   orf102-1     VFNEIPVLLMVAALYLVVFKPFX
                130     140

```

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N.*

*gonorrhoeae*:

```

    orf102.pep  MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL  60
    orf102ng    MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRFMSPL  60
5    orf102.pep  GFGAVVFGAAIPFAAGWGWGSGWVHV KLC LGLMLLAYQLYCGVLLRRFQDYSNAF SHRWYR  120
    orf102ng    GFGAVVFGAAIPFAAGRWGSGWVHV KLC LGLMLLAYQLYCGVLLRRFQDYSNAF SHRWYR  120
10   orf102.pep  VFNEIPVLLMVAALYXVVF KPF  142
    orf102ng    VFNEIPVLLMVAALYL VVF KPF  142

```

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

```

1   ATGATGTTTT CTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCTGTG
15  51  GTTTGCAGGG CTGTTTTACC TGCCGAGGAT TTTCTGCAAT ATGGCGATGA
    101  TTGATGCGCC GCGCGGCAAT CCCGAGTATG TGCGCCTGTC GGGGATGGCG
    151  GTGCGGTTGT ACCGTTTTAT GTCGCCTTTG GGTTCGCGCG CGGTCGTGTT
    201  CGGCGCGGCG ATACCGTTTG CCGCcgggccg GTGGGGCagc ggctggGTTT
    251  ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTATCA GTTGATTATGC
    301  GGCGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
20  351  CTGGTACCGC GTGTTCAACg aATCCCCGT GCTGCTGATG GTTGCCGCGC
    401  TGTATCTGGT CGTGTTCAAA CCGTTTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 764>:

```

1   MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDAPRGN PEYVRLSGMA
25  51  VRLYRFMSPL GFGAVVFGAA IPFAAGRWGS GWVHV KLC LG LMLLAYQLYC
    101  GVLLRRFQDY SNAF SHRWYR VFNEIPVLLM VAALYL VVEK PF*

```

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

```

10      20      30      40      50      60
orf102-1.pep  MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
30  orf102ng    MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRFMSPL
      10      20      30      40      50      60
70      80      90     100     110     120
orf102-1.pep  GFGAVVFGAAIPFAAGWGWGSGWVHV KLC LGLMLLAYQLYCGVLLRRFQDYSNAF SHRWYR
35  orf102ng    GFGAVVFGAAIPFAAGRWGSGWVHV KLC LGLMLLAYQLYCGVLLRRFQDYSNAF SHRWYR
      70      80      90     100     110     120
130     140
orf102-1.pep  VFNEIPVLLMVAALYL VVF KPFX
40  orf102ng    VFNEIPVLLMVAALYL VVF KPFX
      130     140

```

In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

```

45  gi|2314656 (AE000647) conserved hypothetical integral membrane protein
    [Helicobacter pylori] Length = 148
    Score = 79.2 bits (192), Expect = 1e-14
    Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)
50  Query: 3   FSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRFMSPLGF 62
    F W K FH+ VISW A LFYLPRI FV A + V++ +LY F++
    Sbjct: 8   FLWVKAFHVIAVISWMAALFYLPRL FVYHAENAHKKEFVG VVQIQEK--KLYSFIASPAM 65
55  Query: 63  GAVVFGAAIP-----FAAGRWGSGWVHV KLC LGLMLLAYQLYCGVLLRRFQDYSNAF 115
    G + + F+G GW+H KL L ++LLAY YC +R + +
    Sbjct: 66  GFTLITGILMLLIEPTLFKSG----GWLHAKLALVLL LAYHFYCKKCMRELEKDPTRRN 121
60  Query: 116 HRWYRVFNEIPXXXXXXXXXXXXXFKPF 142
    R+YRVFNE P KPF
    Sbjct: 122 ARFYRVFNEAPTILMILIVILVVV KPF 148

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

```

5      1  ATGGCAAAAA  TGATGAAATG  GCGGCTGTT  GCGGCGGTCG  CGGCGGCAGC
      51  GGTTTGGGGC  GGATGGTCTT  AACTGAAGCC  CGAGCCGCAC  GTGCTTGATA
     101  TTACGGAAC  GGTCAAGCGC  GGC // .....
//.. ATTTCTTTA  CGATTTTGTC  CGAACCGGAT  ACGCCGATTA  AGGCGAAGCT
     51  CGACAGCGTC  GACCCCGGGC  TGACCACGAT  GTCGTCGGGC  GGTACAACA
    101  GCAGTACGGA  TACGGCTTCC  AATGCGGTCT  ACTATTATGC  CCGTTCGTTT
     51  GTGCCGAATC  CGGACGGCAA  ACTCGCCACG  GGGATGACGA  CGCAGAATAC
    201  GGTTGAAATC  GACGGCGTGA  AAAATGTGCT  GATTATTCCG  TCGCTGACCG
     51  TGAAAAATCG  CGGCGGCAAG  GCGTTTGTGC  GCGTGTGGG  TCGGACGGC
    301  AAGGCGGCGG  AACGCGAAAT  CCGGACCGGT  ATGAGAGACA  GTATGAATAC
    151  CGAAGTAAA  AGCGGGTTGA  AAGAGGGGGA  CAAAGTGTC  ATCTCCGAAA
     401  TAACCGCCGC  CGAGCAACAG  GAAAGCGGCG  AACGCGCCCT  AGGCGGCCCG
     451  CCGCGCCGAT  AA

```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```

20      1  MAKMMKWAAY  AAVAAAAYWG  GWS.LKPEPH  VLDITETVRR  G.....
      51  .....
     101  .....
     151  .....
     201  ..... I SFTILSEPDT
     251  PIKAKLDSVD  PGLTMMSSGG  YNSSTDASN  AVYYYARFV  PNPDGKLATG
    301  MTTQNTVEID  GVKNVLIIPS  LTVKNRGGKA  FVRVLGADGK  AAEREIRTGM
    351  RDSMNTEVKS  GLKEGDKVVI  SEITAAEQQE  SGERALGGPP  RR*

```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```

30      1  ..GTATCGGTCG  GCGCGCAGGC  ATCGGGGCAG  ATTAAGATAC  TTTATGTCAA
      51  ACTCGGCAA  CAGGTTAAAA  AGGGCGATTT  GATTGCGGAA  ATCAATTCCA
     101  CCTCGCAGAC  CAATACGCTC  AATACGAAA  AATCCAAGTT  GGAAACGTAT
     151  CAGGCGAAGC  TGGTGTGCGC  ACAGATTGCA  TTGGGCGAGC  CGGAGAAGAA
     201  ATATAAGCGT  CAGGCGGCGT  TATGGAAGGA  AAACGCGACT  TCCAAAGAGG
     251  ATTTGGAAAG  CGCGCAGGAT  GCGTTTGCCG  CCGCCAAAGC  CAATGTTGCC
     301  GAGCTGAAGG  CTTTAATCAG  ACAGAGCAA  ATTTCCATCA  ATACCGCCGA
    351  GTCGGAATTG  GGCTACACGC  GCATTACCGC  AACGATGGAC  GGCACGGTGG
     401  TGGCGATTCT  CGTGGAAGAG  GGGCAGACTG  TGAACGCGGC  GCAGTCTACG
     451  CCGACGATTG  TCCAATTGGC  GAATCTGGAT  ATGATGTTGA  ACAAAATGCA
     501  GATTGCCGAG  GGCGATATTA  CCAAGGTGAA  GCGGGGCGAG  GATATTTCTG
     551  TTACGATTTT  GTCCGAACCG  GATACGCCGA  TTAAGGCGAA  GCTCGACAGC
     601  GTCGACCCCG  GGCTGACCAC  GATGTCGTCG  GCGGTTTACA  ACAGCAGTAC
     651  GGATACGGCT  TCCAATGCGG  TCTACTATTA  TGCCCGTTTCG  TTTGTGCCGA
     701  ATCCGGACGG  CAAACTCGCC  ACGGGGATGA  CGACGCAGAA  TACGGTTGAA
     751  ATCGACGGCG  TGA AAAATGT  GCTGATTATT  CCGTCGCTGA  CCGTGAAAAA
     801  TCGCGGCGGC  AAGGCGTTTG  TCGCGGTGTT  GGGTGGCGAC  GGCAAGGCGG
    851  CGGAACGCGA  AATCCGGACC  GGTATGAGAG  ACAGTATGAA  TACCGAAGTA
     901  AAAAGCGGGT  TGAAGAGGG  GGACAAAGTG  GTCATCTCCG  AAATAACCGC
     951  CGCCGAGCAA  CAGGAAAGCG  GCGAACGCGC  CCTAGGCGGC  CCGCCGCGCC
    1001  GATAA

```

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

50      1  ..VSVGAQASGQ  IKILYVKLGQ  QVKKGDLIAE  INSTSQNTNL  NTEKSKLETY
      51  QAKLVSAQIA  LGSAEKKYKR  QAALWKENAT  SKEDLESAQD  AFAAKANVA
     101  ELKALIRQSK  ISINTAESEL  GYTRITATMD  GTVVAILVEE  GQTVNAAQST
     151  PTIVQLANLD  MMLNKMQIAE  GDITKVKAGQ  DISFTILSEP  DTPIKAKLDS
     201  VDPGLTMMSS  GGYNSSTDAT  SNAVYYYARS  FVPNPDGKLA  TGMTTQNTVE
    251  IDGVKNVLI  PSLTVKNRGG  KAFVRVLGAD  GKAAEREIRT  GMRDSMNTEV
    301  KSGLEKGDV  VISEITAAEQ  QESGERALGG  PPRR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

5	orf85.pep	10 20 30 40	MAKMMKWA AVAAVAAAVWGGWS-LKPEPHVLDITETVRRG
	orf85a	10 20 30 40 50 60	MAKMMKWA AVAAVAAAVWGGWSY LKPEPQAAYITETVRRGDISRTVSATGEISPSNLVS
10	orf85.pep	80 90 100	.....ISFTILSEPDTPIKAKLDSVDPGLTTMSSG
	orf85a	210 220 230 240 250 260	TIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSSG
15	orf85.pep	110 120 130 140 150 160	GYNSSTDASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGK
	orf85a	270 280 290 300 310 320	GYNSSTDASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGR
20	orf85.pep	170 180 190 200 210 220	AFVRVLGADGKAAEREIRTGMRDSMNTVEKSGLKEGDKVISEITA AEQQESGERALGGP
	orf85a	330 340 350 360 370 380	AFVRVLGADGKAAEREIRTGMRDSMNTVEKSGLKEGDKVISEITA AEQQESGERALGGP
30	orf85.pep	230	PRRX
	orf85a	390	PRRX

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

35	1	ATGGCAAAAA	TGATGAAATG	GGCGGCTGTT	GCGGCGGTCG	CGGCGGCAGC
	51	GGTTTGGGGC	GGATGCTCTT	ATCTGAAGCC	CGAGCCGCAG	GCTGCTTATA
	101	TTACGGAAAC	GGTCAGGCGC	GGCGACATCA	GCCGGACGGT	TTCTGCAACA
	151	GGGGAGATTT	CGCCGTCCAA	CCTGGTATCG	GTCGGCGCGC	AGGCATCGGG
	201	GCAGATTAAG	AAACTTTATG	TCAAACCTCG	GCAACAGGTT	AAAAAGGCGC
40	251	ATTTGATTGC	GGAAATCAAT	TCGACCTCGC	AGACCAATAC	GCTCAATACG
	301	GAAAAATCCA	AATTGGA AAC	GTATCAGGCG	AAGCTGGTGT	CGGCACAGAT
	351	TGCATTGGGC	AGCGCGGAGA	AGAAATATAA	GCGTCAGGCG	GCGTTGTGGA
	401	AGGATGATGC	GACCGCTAAA	GAAGATTGCG	AAAGCGCACA	GGATGCGCTT
	451	CGCGCGGCCA	AAGCCAATGT	TGCCGAGCTG	AAGGCTCTAA	TCAGACAGAG
45	501	CAAAATTTCC	ATCAATACCG	CCGAGTCGGA	ATTGGGCTAC	ACGCGCATT A
	551	CCGCAACGAT	GGACGGCAGC	GTGGTGGCGA	TTCTCGTGGA	AGAGGGGCAG
	601	ACTGTGAACG	CGGCGCAGTC	TACGCCGACG	ATTGTCCAAT	TGGCGAATCT
	651	GGATATGATG	TTGAACAAAA	TGCAGATTGC	CGAGGGCGAT	ATTACCAAGG
	701	TGAAGGCGGG	GCAGGATATT	TCGTTTACGA	TTTTGTCCGA	ACCGGATACG
50	751	CCGATTAAGG	CGAAGCTCGA	CAGCGTCGAC	CCCGGGCTGA	CCACGATGTC
	801	GTCGGGCGGC	TACAACAGCA	GTACGGATAC	GGCTTCCAAT	GCGGTCTACT
	851	ATTATGCCCG	TTCGTTTGTG	CCGAATCCGG	ACGGCAAAC T	CGCCACGGGG
	901	ATGACGACGC	AGAATACGGT	TGAAATCGAC	GGTGTGAAAA	ATGTGCTGAT
	951	TATTCCGTCG	CTGACCGTGA	AAAATCGCGG	CGGCAGGGCG	TTTGTGCGCG
55	1001	TGTTGGGTGC	AGACGGCAAG	GCGGCGGAAC	GCGAAATCCG	GACCGGTATG
	1051	AGAGACAGTA	TGAATACCGA	AGTAAAAAGC	GGGTGGAAG	AGGGGGACAA
	1101	AGTGGTCATC	TCCGAAATAA	CCGCCGCCGA	GCAGCAGGAA	AGCGGCGAAC
	1151	GCGCCTAGG	CGGCCCGCCG	CGCCGATAA		

This encodes a protein having amino acid sequence <SEQ ID 770>:

60	1	MAKMMKWA AV	AAVAAAVWG	GWSYLKPEPQ	AAYITETVRR	GDISRTVSAT
	51	GEISPSNLVS	VGAQASGQIK	KLYVKLGQQV	KKGD LIAEIN	STSQTNTLNT
	101	EKSKLETYQA	KLVSQAIALG	SAEKKYKRQA	ALWKDDATAK	EDLESAQDAL
	151	AAAKANVAEL	KALIRQSKIS	INTAESELGY	TRITATMDGT	VVAILVEEGQ
	201	TVNAAQSTPT	IVQLANLDMM	LNKMQIAEGD	ITKVKAGQDI	SFTILSEPDT
65	251	PIKAKLDSVD	PGLTTMSSGG	YNSSTDASN	AVYYYARSFV	PNPDGKLATG
	301	MTTQNTVEID	GVKNVLIIPS	LTVKNRGGRA	FVRVLGADGK	AAEREIRTGM



351 RDSMNTEVK SGLKEGDKVVI SEITAAEQQE SGERALGGPP RR\*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

5	orf85a.pep	30 40 50 60 70 80	PQAA YITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE
	orf85-1		VSVGAQASGQIKKLYVKLGQQVKKGDLIAE
10	orf85a.pep	90 100 110 120 130 140	INSTSQTNLTNTEKSKLETYQAKLVSAQIALGSAEKYKQQAALWKDDATAKEDLESAQD
	orf85-1		INSTSQTNLTNTEKSKLETYQAKLVSAQIALGSAEKYKQQAALWKENATSKEDLESAQD
15	orf85a.pep	150 160 170 180 190 200	ALAAKANKVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
	orf85-1		AFAAAKANKVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
20	orf85a.pep	210 220 230 240 250 260	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
	orf85-1		PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
25	orf85a.pep	270 280 290 300 310 320	GGYNSSTD TASNAVYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG
	orf85-1		GGYNSSTD TASNAVYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG
30	orf85a.pep	330 340 350 360 370 380	RAFVRVLGADGKAAEREIRTGMRDSMNTEVK SGLKEGDKVVI SEITAAEQQESGERALGG
	orf85-1		KAFVRVLGADGKAAEREIRTGMRDSMNTEVK SGLKEGDKVVI SEITAAEQQESGERALGG
35	orf85a.pep	390	P P R R X
	orf85-1		P P R R X

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

#### Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

50	ORF85	1	MAKMMKWAAVA AAAAAVWGWS.LKPEPHVLDITETVRRG.....	40
	ORF85ng	1	MAKMMKWAAVA AAAAAVWGWSY LKPEPQAAYITEAVRRGDISRTVSAT	50
55	ORF85	201	TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDT	250
	ORF85ng	201	TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDT	250
60	ORF85	251	PIKAKLDSVDPGLTTMSSGGYNSSTD TASNAVYYARSFVNPDPGKLATG	300
	ORF85ng	251	PIKAKLDSVDPGLTTMSSGGYNSSTD TASNAVYYARSFVNPDPGKLATG	300
65	ORF85	301	MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM	350
	ORF85ng	301	MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM	350
	ORF85	152	RDSMNTEVK SGLKEGDKVVI SEITAAEQQESGERALGGPPRR	393
	ORF85ng	351	KDSMNTEVK SGLKEGDKVVI SEITAAEQQESGERALGGPPRR	393

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```

1  ATGGCAAAAA TGATGAAATG GCGGGCTGTT GCGGCGGTCG CCGCGGCaac
51  GGTTTGGGGC GGATGGTCTT ATCTGAAGCC CGAACCCGAG GCTGCTTATA
101 TTAGCGAaac ggTCAGGCGC GGCGATATCA GCCGGACGGT TTCCGCGACG
151 GcgAGATT T CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCTTCGGG
201 GCAGATTAAA AAGCTTTATG TCAAACTCGG GCAACAGGTC AAAAAGGGCG
251 ATTTGATTGC GGAAATCAAT TCGACCACGC AGACCAACAC GATCGATATG
301 GAAAAATCCA AATTGGAAAC GTATCAGGCG AAGCTGGTGT CGGCACAGAT
351 TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTTGTGGA
401 AGGATGATGC GACCTCTAAA GAAGATTTGG AAAGCGCGCA GGATGCGCTT
451 CCGCGCGCCA AAGCCAATGT TGCCGAGTTG AAGGCTTTAA TCAGACAGAG
501 CAAAATTTCC ATCAATACCG CCGAGTCGGA TTTGGGCTAC ACGCGCATT
551 CCGCGACGAT GGACGCGACG GTGGTGGCGA TTCCCGTGA AGAGGGGCG
601 ACTGTGAACG CCGCGCAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
651 GGATATGATG TTGAACAAA TGCAGATTGC CGAGGCGCAT ATTACCAAGG
701 TGAAGCGGG GCAGGATATT TCGTTTACGA TTTTGTCCGA ACCGGATACG
751 CCGATTAAGG CGAAGCTCGA CAGCGTCGAC CCGGGGCTGA CCACGATGTC
801 GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTATT
851 ATTATGCCCC TTCGTTTGTG CCGAATCCGG ACGGCAAACT CGCCACGGGG
901 ATGACGACGC AGAATACGGT TGAAATCGAC GGTGTGAAAA ATGTGTTGCT
951 TATTCGCTCG CTGACCGTGA AAAATCGCGG CGCAAGGCG TTCGTACGCG
1001 TGTTGGGTGC GGACGGCAAG GCAGTGGAA CCGAAATCCG GACCGGTATG
1051 AAAGACAGTA TGAATACCGA AGTGAAAGC GGTTTGAAAG AGGGGGACAA
1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC
1151 GCGCCCTAGG CCGCCGCCCG GCCTGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```

1  MAKMMKWA AV AAVAAAAVWG GWSYLKPEPQ AAYITEAVRR GDISRTVSAT
51  GEISPSNLVS VGAQASGQIK KLYVKLGQOV KKGDLIAEIN STTQNTIDM
101 EKSKLETYQA KLVSAQIALG SAEKKYKRQA ALWKDDATSK EDLESAQDAL
151 AAKANVAEL KALIRQSKIS INTAESDLGY TRITATMDGT VVAIPVEEGQ
201 TVNAAQSTPT IVQLANLDM LNKMQIAEGD ITKVKAGQDI SFTILSEPDT
251 PIKAKLDSVD PGLTTMSSG YNSSTDASN AVYYYARFV PNPDKLATG
301 MTQNTVEID GVKNVLLIPS LTVKNRGGKA FVRVLGADGK AVEREIRTGM
351 KDSMNTVEKS GLKEGDKVVI SEITAEQQE SGERALGGPP RR*

```

ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

```

30      40      50      60      70      80
orf85ng  PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE
40      orf85-1  VSVGAQASGQIKILYVKLGQQVKKGDLIAE
                        10      20      30

90      100     110     120     130     140
orf85ng  INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEDLESAQD
45      orf85-1  INSTSQNTLNTTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKENATSKEDLESAQD
                        40      50      60      70      80      90

150     160     170     180     190     200
orf85ng  ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVVAIPVEEGQTVNAAQST
50      orf85-1  AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
                        100     110     120     130     140     150

210     220     230     240     250     260
orf85ng  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPLTTMSS
55      orf85-1  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPLTTMSS
                        160     170     180     190     200     210

270     280     290     300     310     320
orf85ng  GGYNSSTDASNNAVYYYARFVNPDPGLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
60      orf85-1  GGYNSSTDASNNAVYYYARFVNPDPGLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
                        220     230     240     250     260     270

330     340     350     360     370     380

```

-431-

```

5      orf85ng      KAFVRVLGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVIS EITAAEQQESGERALGG
      orf85-1      KAFVRVLGADGKAAEREIRTGMKDSMNTEVKSGLKEGDKVVIS EITAAEQQESGERALGG
                                280      290      300      310      320      330

      orf85ng      390
      orf85ng      PPRRX
      orf85-1      PPRRX

```

10 In addition, ORF85ng shows significant homology to an *E.coli* membrane fusion protein:

```

      gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from
      membrane fusion protein precursor, MTRC_NEIGO SW: P43505 (412 aa) [Escherichia
      coli] Length = 380
      Score = 193 bits (485), Expect = 2e-48
15      Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

      Query: 29  PQAA YITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGLIAE 88
      P   Y T  VR GD+ ++V ATG++      V VGAQ SGQ+K L V +G +VKK L+
      Sbjct: 41  PVPTYQTLIVRPGDLQQSVLATGKLDALRKVDVGAQVSGQLKTL SVAIGDKVKKDQLLGV 100

20      Query: 89  INSTTQTNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEXXXXXXX 148
      I+   N I  ++ L  +A+  A+  L  A  Y RQ  L  +  A S++
      Sbjct: 101 IDPEQAENQIKEVEATLMELRAQRQQAELKLARVTYSRQRLAQTKAVSQDLDTAAT 160

25      Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST 208
      I++++ S++TA+++L YTRI A M G V I  +GQTV AAQ
      Sbjct: 161 EMAVKQAQIGTIDAQIKRNQASLDTAKTNLDYTRIVAPMAGEVTQITTLQGQTVIAAQQ 220

30      Query: 209 PTIVQLANLDMMLNKMQIAEGDITKVKGQDISFTILSEPDTPIKAKLDSVPDGLTTMSS 268
      P I+ LA++ ML K Q++E D+  +K GQ  FT+L +P T  + ++ V P
      Sbjct: 221 PNILTLADMSAMLVKAQVSEADVHLKPGQKAWFTVLGDPLTRYEGQIKDVLP----- 273

      Query: 269 GGYNSSTDASNAVYYIARSFVPNPDGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG 328
      + +  ++A++YYAR VNP+G L  MT Q  +++ VKNVL IP  + + G
35      Sbjct: 274 -----TPEKVNDIAIFYARFEVVPNENGLRLDMTAQVHIQLTDVKNVLTIPLSALGDPVG 328

      Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVIS E 372
      +V L  +G+  ERE+  G ++  + E+  GL+ GD+VVI E
      Sbjct: 329 DNRYKVKLLRNGETREREV TIGARNDTDVEIVKGLEAGDEVVIGE 373

```

40 Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein  
45 was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 92

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 773>:

```

50      1  ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAATCGT
      51  TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
      101  CGGTTGTCGG CAATACCCTG CACCCTACCT ACTATAGAGA CATACGCAGG
      151  GGCAAACGTG ATGCGGAAGc CAAATTCGCC GACgGcAGCG TAACTTACGG
      201  CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAAGGCT ATGGATTGT

```

-432-

251 TCACGCTTGC CTGGCAGTTG GCGGCAAATG ACGCGAAACT CCCCCGGGG  
 301 CTGAAAATCA CCAACGGCAA AAACTTTAT TCCGTCGGCG GTTTGAATAA  
 351 GGCGGGTACA GGAAAATACA GCATAGGCGG CGTGGAACC GAAGTCGTCA  
 401 AATATCGGGT GCGGCGCGG GACGATGCGG TAATGTATTT cTTGCGACCG  
 5 TCCCTGAACA ATATTCCGGC ACAAATCGGC TATACCGACG ACGGCAAAAC  
 501 CTATACGCTG AAACCTCAAT CCGTGCAGAT CAACGGCCAG GCAGCCAAAC  
 551 CGTAA

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

1 ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNTL HPTYYRDIRR  
 10 51 GKLYAEAKFA DGSVTYKGAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG  
 101 LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP  
 151 SLNNIPAQIG YTDDGKTYTL KLKSVQINGQ AAKP\*

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

1 ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC  
 15 51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT  
 101 ATTCGGGCAG CTACGGCATT CCGGCCACGA TGACATTTGA ACGCAGCGGC  
 151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG  
 201 TTTCGAGTCC GCGGTACGG TTGTGCGCAA TACCCTGCAC CCTACCTACT  
 251 ATAGAGACAT ACGCAGGGGC AAACCTGTATG CGGAAGCCAA ATTCGCCGAC  
 20 301 GGCAGCGTAA CTTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAGCCC  
 351 CAAGGCTATG GATTTGTTCA CGCTTGCCTG GCAGTTGGCG GCAAATGACG  
 401 CGAAACTCCC CCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC  
 451 GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT  
 501 GGAAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA  
 25 551 TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT  
 601 ACCGACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA  
 651 CGGCCAGGCA GCCAAACCGT AA

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

1 MMKTFKNIFS AAILSALPC AYAAGLPQSA VLHYSGSYGI PATMTFERSG  
 30 51 NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD  
 101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS  
 151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY  
 201 TDDGKTYTLK LKSVQINGQA AKP\*

Computer analysis of this amino acid sequence gave the following results:

### 35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N. meningitidis*:

				10	20	30
40	orf120.pep			IPATMTFERSGNAYKIVSTIKVPLYNIRFE		
	orf120a	SAAILSALPCAYAAGLPXSAVLHYSGSYGIPATXXXXXXXXNAXKIVSTIKVPLYNIRFE				
		10 20 30 40 50 60				
45	orf120.pep	SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYKGAGESKTEQSPKAMDFTLAWQL				
	orf120a	SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAXXXXXXQSPKAMDFTLAWQL				
		70 80 90 100 110 120				
50	orf120.pep	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE EVVKYRVRRGDDAVMYFFAP				
	orf120a	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE EVVKYRVRRGDDAVMYFFAP				
55		130 140 150 160 170 180				

-433-

```

              160      170      180
orf120.pep  SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAPX
              |||
orf120a     SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAPX
5           190      200      210      220

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

1  ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CNAATCCGCC GTGCTGCACT
101 ATTCGGGCAG CTACGGCATT CCGGCCACNA NNANNTNNGN ACNNNGNGNC
151 AATGCTTNCA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
201 TTTCGAGTCC GCGGTACGG TTGTCGGCAA TACCCTGCAC CCTACCTACT
251 ATAGAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCGGAC
301 GGCAGCGTAA CCTACGGCAA AGCGGNNNNN ANCNNNNNNG NGCAAAGCCC
351 CAAGGCTATG GATTTGTTCA CGCTTGCNTG GCAGTTGGCG GCAAATGACG
15 401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTC
451 GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT
501 GGAAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA
551 TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
20 601 ACCGACGACG GCAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
651 CGGCCAGGCA GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

1  MMKTFKNIFS AAILSAALPC AYAAGLPXSA VLHYSGSYGI PATXXXXXXXX
51 NAXKIVSTIK VPLYNIRFES GGTVVGNLTH PTYYRDIRRG KLYAEAKFAD
101 GSVTYGKAXX XXXXQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
25 151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAIQIGY
201 TDDGKTYTLK LKSVQINGQA AKP*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

              10      20      30      40      50      60
orf120a.pep  MMKTFKNIFS AAILSAALPCAYAAGLPXSAVLHYSGSYGIPATXXXXXXXXNAXKIVSTIK
30          |||
orf120-1     MMKTFKNIFS AAILSAALPCAYAAGLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf120a.pep  VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYGKAXXXXXXQSPKAM
35          |||
orf120-1     VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
              70      80      90      100     110     120

              130     140     150     160     170     180
orf120a.pep  DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
40          |||
orf120-1     DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
              130     140     150     160     170     180

              190     200     210     220
orf120a.pep  DAVMYFFAPSLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAPX
50          |||
orf120-1     DAVMYFFAPSLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAPX
              190     200     210     220

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

```

55  orf120.pep  IPATMTFERSGNAYKIVSTIKVPLYNIRFE 30
      |||
orf120ng  SAILSAALPCAYAARLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIKVPLYNIRFE 69

orf120.pep  SGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMD LFTLAWQL 90
60  |||
orf120ng  SGGTVVGNLHPAYYKDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMD LFTLAWQL 129

```

-434-

5 orf120.pep AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP 150  
 orf120ng AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDTVTYFFAP 189  
 orf120.pep SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP 184  
 orf120ng SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP 223

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

10 1 ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC  
 51 CCTGCCGTGC GCGTATGCGG CAAGGCTACC CCAATCCGCC GTGCTGCACT  
 101 ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTGTA ACGCAGCGGC  
 151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG  
 201 TTTCGAATCC GCGCGTACGG TTGTCCGCAA TACCCTGCAC CCTGCCTACT  
 15 251 ATAAAGACAT ACGCAGGGGC AAACGTATG CGGAAGCCAA ATTCGCCGAC  
 301 GGCAGCGTAA CCTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC  
 351 CAAGGCTATG GATTGTTC ACGTTGCCGT GCAGTTGGCG GCAAATGACG  
 401 CGAAACTCCC CCCGGGTCTG AAAATCACCA ACGGCAGAAA ACTTTATTCC  
 451 GTCGGCGGCC TGAATAAGGC GGTACGGGA AAATACAGCA TaggCGGCGT  
 20 501 GGAAACCGAA GTCGTCAAT ATCGGGTGCG GCGCGGCGAC GATACGGTAA  
 551 CGTATTCTT CGACCGTCC CTGAACAATA TTCGGGCACA AATCGGCTAT  
 601 ACCGACGACG GCAAAACCTA TACGCTGAAG CTCAATCGG TGCAGATCAA  
 651 CGGACAGGCC GCCAAACCGT AA

This encodes a protein having amino acid sequence <SEQ ID 780>:

25 1 MMKTFKNIFS AAILSALPC AYAARLPQSA VLHYSYSGYGI PATMTFERSG  
 51 NAYKIVSTIK VPLYNIRFES GGTVVGNLTH PAYYKDIRRG KLYAEAKFAD  
 101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS  
 151 VVGLNKAGTG KYSIGGVETE VVKYRVRRGD DTVTYFFAPS LNNIPAQIGY  
 201 TDDGKTYTLK LKSVQINGQA AKP\*

30 In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

35 orf120-1.pep 10 20 30 40 50 60  
 orf120ng MMKTFKNIFS AAILSALPC AYAAGLPQSA VLHYSYSGYGI PATMTFERSG NAYKIVSTIK  
 10 20 30 40 50 60  
 40 orf120-1.pep 70 80 90 100 110 120  
 orf120ng VPLYNIRFES GGTVVGNLTH PYYKDIRRG KLYAEAKFAD GSVTYGKAGESKTEQSPKAM  
 70 80 90 100 110 120  
 45 orf120-1.pep 130 140 150 160 170 180  
 orf120ng DLFTLAWQLA ANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD  
 130 140 150 160 170 180  
 50 orf120-1.pep 190 200 210 220  
 orf120ng DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX  
 190 200 210 220

55 This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

-435-

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC  
 51 .GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA  
 101 CTCCGTTTGC GGTTCGGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC  
 151 GAATGGTTGC AGAAAAAGGG TTGAACCGT GCATCCGCTT CGATGTCTGT  
 201 GATGGTGTTC TCCTTGATTT TGTGTTGGC ATTATTGTTG ATTATCGTCC  
 251 CTATGCTGCT CGGGCAGTTC AACAAATTTGG CATCGCGCCT GCCCAATTA  
 301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 351 CCGATATCTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC  
 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG  
 451 AGGCAGGGCG GCAATATT..

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

1 MYRRKGRGIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL  
 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 151 RQGGNI..

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC  
 51 GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA  
 101 CTCCGTTTGC GGTTCGGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC  
 151 GAATGGTTGC AGAAAAAGGG TTGAACCGT GCATCCGCTT CGATGTCTGT  
 201 GATGGTGTTC TCCTTGATTT TGTGTTGGC ATTATTGTTG ATTATCGTCC  
 251 CTATGCTGCT CGGGCAGTTC AACAAATTTGG CATCGCGCCT GCCCAATTA  
 301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 351 CCGATATCTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC  
 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG  
 451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCTTCC  
 501 CTTGCTGCTT TACTATTTC TGTGGATTG GCAGCGGTGG TCGTGCGGCA  
 551 TTGCCAAACT GGTTCGAGG CGTTTGGCCG GTGCTTATAC GCGCATTACA  
 601 GGCAATTTGA ACGAGGTATT GGGCGAATTT TTGCGCGGGC AGCTTCTGGT  
 651 AATGCTGATT ATGGGCTTGG TTACGGTTT GGGATTGGTG CTGGTCGGGC  
 701 TGGATTCGGG GTTTCGCATC GGTATGCTTG CCGGTATTTT GGTGTTTGTC  
 751 CTTATCTCG GGGCGTTTAC GGGATTGCTG CTTGCCACCG TCGCCGCTT  
 801 GCTCCAGTTC GGTTCGTGGA ACGGCATCCT ATCGGTTTGG GCGGTTTTTG  
 851 CCGTAGGACA GTTCTCGAA AGTTTTTCA TTACCCGAA AATCGTGGA  
 901 GACCGTATCG GGCTGTCGCC GTTTTGGGT ATCTTTTCGC TGATGGCGTT  
 951 CGGGCAGCTG ATGGGCTTTC TCGGAATGTT GCGGGGATTG CCTTTGCCG  
 1001 CCGTAACCTT GCTCTGCTT CCGGAGGGCG TGCAGAAATA TTTTGCCGGC  
 1051 AGTTTTTACC GGGCAGGTA G

This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

1 MYRRKGRGIK PWMGAGAAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL  
 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 151 RQGGNIVSSI GNLLLLPLLL YYFLLDWQRW SCGIKLVPR RFAGAYTRIT  
 201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLV LVGLDSGFAI GMLAGILVFE  
 251 PYLGAFTGLL LATVAALLQF GSWNGILSVW AVFAVGQFLE SFFITPKIVG  
 301 DRIGLSPFVW IFSLMAFGQL MGFVGMLAGL PLAAVTLVLL REGVQKYFAG  
 351 SFYRGR\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N. meningitidis*:

10 20 30 40 50 60  
 orf121.pep MYRRKGRGIKPWMGAGXAFAALVWLVFALGDTLTPFAVAALVAYVLDPLVEWLQKKGLNR  
 55 orf121a MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAALVAYVLDPLVEWLQKKGLNR  
 10 20 30 40 50 60  
 70 80 90 100 110 120  
 orf121.pep ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV

-436-

```

      |||
orf121a  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
              70      80      90      100      110      120

5      130      140      150
orf121.ppep  EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNI
      |||
orf121a  EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
              130      140      150      160      170      180

10      SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGF
              190      200      210      220      230      240

```

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

```

15      1  ATGTATCGGA  GGAAAGGGCG  GGGCATCAAG  CCGTGGATGG  ATGCCGGTGC
      51  GGCGTTTGCC  GCCTTGCTCT  GGCTGGTTT  CGCGCTCGGC  GATACTTTGA
      101  CTCCGTTTGC  GGTTCGGCG  GTGCTGGCGT  ATGTATTGGA  CCCTTTGGTC
      151  GAATGGTTGC  AGAAAAAGGG  TTGAACCGT  GCATCCGCTT  CGATGCTGTG
      201  GATGGTGTTC  TCCTTGATT  TGTGTGGC  ATTATTGTG  ATTATTGTCC
      251  CTATGCTGGT  CGGGCAGTTC  AACAATTTGG  CATCGCGCCT  GCCCAATTA
      301  ATCGGTTTTA  TGCAGAACAC  GCTGCTGCCG  TGGTTGAAA  ATACAATCGG
      351  CCGATATGTC  GAAATCGATC  AGGCATCTAT  TATTGCGTGG  CTTCAGGCGC
      401  ATACGGGCGA  GTTGAAGAAC  GCGCTTAAG  CGTGGTTTCC  CGTTTTGATG
      451  AGGCAGGGCG  GCAATATTGT  CAGCAGTATC  GGCAACCTGC  TGCTGCTTCC
      501  CTTGCTGCTT  TACTATTTC  TGCTGGATTG  GCAGCGGTGG  TCGTGCGGCA
      551  TTGCCAACT  GGTTCGAGG  CGTTTGGCG  GTGCTTATAC  GCGCATTACA
      601  GGCAATTGTA  ACGAGGTATT  GGGCGAATT  TTGCGCGGGC  AGCTTCTGGT
      651  GATGCTGATT  ATGGGTTTG  TTTACGGCT  GGGGTTGGTG  CTGGTCGGGC
      701  TGGATTCGGG  GTTTCGAATC  GGTATGGTTG  CCGGTATTTT  GGTTTTGTG
      751  CCCTATTGG  GCGCGTTTAC  AGGACTGCTG  CTGGCAACCG  TCGCCGCCTT
      801  GCTCCAGTTC  GGTTCGTGGA  ACGGCATCT  GGCTGTTTGG  GCGGTTTTTG
      851  CCGTAGGACA  GTTCTCGAA  AGTTTTTCA  TTACGCCGAA  AATCGTGGGA
      901  GACCGTATCG  GCCTGTCGCC  GTTTTGGGT  ATCTTTTCGC  TGATGGCGTT
      951  CGGGCAGCTG  ATGGGCTTTG  TCGGAATGTT  GGCCGATTG  CCTTTGCGCG
      1001  CCGTAACCTT  GGTCTTGCTT  CCGGAGGGCG  TGCAGAAATA  TTTTGCCGGC
      1051  AGTTTTTACC  GGGGCAGGTA  G

```

This encodes a protein having amino acid sequence <SEQ ID 786>:

```

      1  MYRRKGRGIK  PWDAGAAFA  ALVWLVFALG  DLTLPFAVAA  VLAYVLDPLV
      51  EWLQKKGLNR  ASASMSVMVF  SLILLALL  IIVPMLVGQF  NNLASRLPQL
      101  IGFMQNTLLP  WLKNTIGGYV  EIDQASIIAW  LQAHTGELSN  ALKAWFPVLM
      151  RQGGNIVSSI  GNLLLLPLL  YYFLLDWQRW  SCGIAKLVPR  RFAGAYTRIT
      201  GNLNEVLGEF  LRQQLLVMLI  MGLVYGLGLV  LVGLDSGF  GMVAGILVFE
      251  PYLGAFTGLL  LATVAALLQF  GSWNGILAVW  AVFAVGQFLE  SFFITPKIVG
      301  DRIGLSPFWV  IFSLMAFGQL  MGFVGMLAGL  PLAAVTLVLL  REGVQKYFAG
      351  SFYRGR*

```

ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

```

      10      20      30      40      50      60
orf121a.ppep  MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
      |||
orf121-1  MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
      10      20      30      40      50      60

      70      80      90      100      110      120
orf121a.ppep  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      |||
orf121-1  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
      70      80      90      100      110      120

      130      140      150      160      170      180
orf121a.ppep  EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
      |||
orf121-1  EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
      130      140      150      160      170      180

      190      200      210      220      230      240
orf121a.ppep  SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGF

```



40	1	ATGTATCGGA	GAAAAGGACG	GGGCATCAAG	CCGTGGATGG	GTGCCGCGCG
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTA	CGCGCTCGGC	GATACTTTGA
	101	CTCCGTTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTGTTGGA	CCCTTTGGTC
	151	GAATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCAATCCGCTT	CATATGTCGT
45	201	GATGCTGTTT	TCCTTGATT	TGTTGTTGGC	ATTTATGTTG	GGATGTGTCC
	251	CTATGCTGGT	CGGGCAGTTC	AATAATTTGG	CATCTCGCCT	GCCCCAATTA
	301	ATCGTTTTTA	TGCAGAACAC	GCTGCTGCCG	TGGTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGCGTGG	TTTCAGGCGC
50	401	ATACGGGCGA	GTTAGCAAC	CGGCTTAAGG	CGTGTTTCC	CGTTTGTATG
	451	AAACAGGGCG	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCCGCC
	501	CTTGCTGCTT	TACTATTTCC	TGCTGGATTG	GCAGCGGTGG	TCGTGCCGCA
	551	TCGCCAAACT	GGTTCCGAGG	CGTTTTGCCG	GTGCTTATAC	GCGCATTACG
55	601	GGTAATTTGA	ACGAGGTATT	GGCGCAATTT	TTGCGCGGTC	AGCTTCTGGT
	651	GATGCTGATT	ATGGGCTTGG	TTTACGGTTT	GGGATTGATG	CTAGTCGGAC
	701	TGGATTCCGG	ATTTGCCATC	GGTATGGTTG	CCGGTATTTT	GGTGTTTGTC
	751	CCCTATTTGG	GTGCGTTTAC	GGGATTGCTG	CTTGCCACTG	TTTGACGCTT
60	801	GCTCCAGTTC	GGTTCGTGGA	ACGGAATCTT	GGCTGTTTGG	CGCGTTTTTG
	851	CCGTCGGTCA	GTTTCTCGAA	AGTTTTTTCA	TTAGCCCGAA	AATTGTAGGA
	901	GACCGTATCG	GCCTGTCGCC	GTTTTGGGTT	ATCTTTTCGC	TGATGGCGTT
	951	CGGAGAGCTG	ATGGGGCTTG	TCGGAATGTT	GGCCGGATTG	CCTTTGCCCG
60	1001	CCGTAACCTT	GGTCTTGCTT	CGCGAGGGCG	CGCAGAAATA	TTTTGCCGGC
	1051	AGTTTTTACC	GGGCCAGGTA	G		

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

```

1 MYRRKGRGIK PWMGAGAAFA ALVWLIVYALG DTLTFFAVAA VLAYVLDPLV
51 EWLQKKGKLN ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM
151 KQGGNIVSSI GNLLLPPLLL YYFLLDWQRW SCGIAKLVPR RFAGAYTRIT
201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLM LVGLDSGFAI GMVAGILVVF
251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
301 DRIGLSPFWV IFSLMAFGEL MGFVGMLAGL PLAAVTLVLL REGAQKYFAG
351 SFYRGR*

```

10 ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:

```

15 orf121-1.pep      10      20      30      40      50      60
    MYRRKGRGIK PWMGAGAAFAALVWLVFALGDTLTFFAVAAVLAYVLDPLVEWLQKKGKLN
orf121ng-1          10      20      30      40      50      60
    MYRRKGRGIK PWMGAGAAFAALVWLIVYALGDTLTFFAVAAVLAYVLDPLVEWLQKKGKLN

20 orf121-1.pep      70      80      90      100     110     120
    ASASMSVMVFS LILLALLL IIVPMLVGQFN NNLASRLPQLIGFMQNTLLPWLKNTIGGYV
orf121ng-1          70      80      90      100     110     120
    ASASMSVMVFS LILLALLL IIVPMLVGQFN NNLASRLPQLIGFMQNTLLPWLKNTIGGYV

25 orf121-1.pep      130     140     150     160     170     180
    EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLP LLLYYFLLDWQRW
orf121ng-1          130     140     150     160     170     180
    EIDQASIIAWFQAHTGELSNALKAWFPVLMKQGGNIVSSIGNLLLP LLLYYFLLDWQRW

30 orf121-1.pep      190     200     210     220     230     240
    SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVVLGLDSGFAI
orf121ng-1          190     200     210     220     230     240
    SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLMLVGLDSGFAI

35 orf121-1.pep      250     260     270     280     290     300
    GMLAGILVFPY LGAFTGLL LATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG
orf121ng-1          250     260     270     280     290     300
    GMVAGILVFPY LGAFTGLL LATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG

40 orf121-1.pep      310     320     330     340     350
    DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX
orf121ng-1          310     320     330     340     350
    DRIGLSPFWVIFSLMAFGELMGFVGMLAGLPLAAVTLVLLREGAQKYFAGSFYRGRX

```

In addition, ORF121ng-1 shows homology to a permease from *H. influenzae*:

```

50 sp|P43969|PERM_HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
   Score = 69.9 bits (168), Expect = 2e-11
   Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)

Query: 26 VYALGDTLTFFAVAAVLAYVLDPLVEWL-QKKGKLNASASMSVMVFSXXXXXXXXXXVP 84
      +Y GD + P +A VL+Y+L+ + +L Q R A++ + VP
Sbjct: 32 IYFPGDLIAPLLIALVLSYLLEIPINFLNQYLKCPRLATILIFGSFIGLAAVFFLVLP 91

55 Query: 85 MLVGQFN NNLASRLPQLIGFMQNTLLPWLKNTIGGYVE-IDQASIIAWFQAHTGELSNALK 143
      ML Q +L S LP + N WL N Y E ID + + + F + ++ +
Sbjct: 92 MLWNQITISLSDLPAMF----NKSNEWLLNLPKNYPELIDYSMVDSIFNSVREKILGFE 147
Query: 144 AWFVLMKQGGNIVSSIGNXXXXXXXXXXXXXDWQRWSCGIAKLVPRRFAGAYTRITGNL 203

60 + + + N+VS D G+++ +P+ A+ R +
Sbjct: 148 SAVKLSLASIMNLVSLGIYAFLVPLMMFFMLKDKSELLQGVSRFLPKNRNLAFXRWK-EM 206

Query: 204 NEVLGEFLRGQXXXXXXXXXXXXXXXXXXXXSGFAIGMVAGILVFPYXXXXXXXXXXXX 263
      + + ++ G+ + + G+ V VPY
65 Sbjct: 207 QQQISNYIHGKLEILIVTLITYIIFLIFGLNYP LLLAFAVGLSVLVPIYIGAVIVTIPVA 266

```

Query: 264 XXXXXQFGSWNGILAVWAVFAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323  
 QFG + FAV Q L+ + P + + + L P +I S++ FG L GF  
 Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIIIISVLIFGGLWGF 326

5 Query: 324 VGMLAGLPLAAVTLVLL 340  
 G+ +PLA + ++  
 Sbjct: 327 WGVFFAIPLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and  
 10 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

15 1 ..ACTGCTTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT  
 51 TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT  
 101 TTTGCACGTC CTGCCCCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC  
 151 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG  
 201 TTTTGCTTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG  
 251 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTTCTGTGC  
 20 301 AATCACGGTC GTATCGACAT TGACCGCTG CCAACCCTGC GCCTGAACGC  
 351 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT  
 401 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
 451 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC  
 501 CGAGCAGCCC TTTTTCAAAT GGGATTTCOA CTCCGCCAAG TATCAG..

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

1 ..TAFSAALRLS PSXLVIFLSF GKPYYQQTAAI LTFECTSCPP RSNAYQQYRR  
 51 LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRRECGLFC  
 101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCTRT  
 151 EQRVGNVQVQ RIGIGVSEQP FFKWDFNSAK YQ..

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

1 ATATCGTACT GGGCAAGCAG TTCGCCGAT TTTTGAAG TAGATACCGC  
 51 GCCTTTGATT TTTTGGCGC TCTTACCCAA GGCTTCGATG AAAAAGTTGA  
 101 TGGTCGAGCC GGTACCGATG CCGATATATT CATTTTCGGG TACGAATTTCG  
 35 151 ACTGCTTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTGTG TCGTCATATT  
 201 TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT  
 251 TTTGCACGTC CTGCCCCGCG CGTTCAAATG CGTACCAGCA ATACCGCCGC  
 301 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG  
 351 TTTTGCTTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG  
 40 401 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTTTGG GTTTCTGTGC  
 451 AATCACGGTC GTATCGACAT TGACCGCTG CCAACCCTGC GCCTGAACGC  
 501 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT  
 551 GCGGCGGTGT CGGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
 601 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC  
 651 CGAGCAGCCC TTTTTCAAAT GGGATTTCOA CTCCGCCAAG TATCAGCTTT  
 45 701 CTGCCTTCGG TCAATTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT  
 751 CGTCATCGTT TGTGTTCTCG A

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

1 ISYWASSSPD FLEVDTAPLI FLPLLPKASM KKLMEVPVPM PIYSFSGTNS  
 50 51 TAFSAAMRLS SSCVVIFLSF GKPYYQQTAAI LTFECTSCPP RSNAYQQYRR  
 101 LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRREFGLFC  
 151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCTRT  
 201 EQRVGNVQVQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSPTDV  
 251 RHRLCS\*

Computer analysis of this amino acid sequence gave the following results:

5	orf122.pep						10	20	30
						TAFSAALRLSPSXLVIFLSFGKPYQQTAAI			
						:       :			
	orf122a	FLPLLPKASMKKLMVEPVMPMYSFSGTNSTAFSAAMRLSSSCVVIFLSFGKPYQQTAAI							
		30	40	50	60	70	80		
10		40	50	60	70	80	90		
	orf122.pep	LTFFCTSCPPRSNAYQYYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR							
	orf122a	LTFFXTSCPPRSNPYQYYRRLRLYAFHAPEITEFFVGFAFXVDARNVYAQIGGDVGTHLR							
		90	100	110	120	130	140		
15		100	110	120	130	140	150		
	orf122.pep	NVRREGFLCNHGRIDIDLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQT CRT							
		:							
	orf122a	NMRREFGLCNHGRIDIDLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQT CRT							
		150	160	170	180	190	200		
		160	170	180					
	orf122.pep	EQRVGNVGQQRIGIGVSEQPFFKWDFNSAKYQ							
25	orf122a	EQRVGNVGQQRIGIGVSEQPFFKWDFNSAKYQLSAFGQLVDIVALS DTDVHRRLCSX							
		210	220	230	240	250			

	1	ATATCATATT	GGGCAAGCAG	TCTACTGGAT	TTTTTGGAAG	TAGATACCGC
30	51	GCCTTTGATT	TTTTTGCCGC	TCTTACCCAA	GGCTTCGATG	AAAAAGTTGA
	101	TGGTCGAATC	GGTACCGATG	CCGATGTATT	CGTTTTCGGG	TACGAATTCCG
	151	ACTGCTNTTT	CGCGCGGATC	CGCGTTGAGT	TCGCTTTGTG	TCGTCATATT
	201	TTTGTCTTTT	GGGAAACCGT	ATCAACAAAC	AGCCGCGATC	TTAACATTTT
35	251	TTNNNACGTC	CTGCCCCCGC	CGTTCAAATC	CTTACCAGCA	ATACCGCCGC
	301	CTGCGACTCT	ATGCCCTTCCA	TGCGCCCGAG	ATAACCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GANGTTTGACG	CACGAAATGT	CTATGCCCAA	ATCGGCCGGC
	401	ATGTTGGCAC	GCATTTGCGG	AATATGCGGC	GCGAGTTTGG	GTTTCTGTGC
40	451	AATCACGGTC	GTATCGACAT	TGACCGCCTG	CCAACCTCTG	GCCTGAACGC
	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCCGATC	TTTGAACCTT
	551	GCGGCGGTGT	CGCGGAAATG	GCTGCCGATA	TCGCCCAAAC	CTGCCGCCTC
	601	GAGCAGCGCG	TCGGTAACGG	CGTGCAGCAG	CGCATCGGCA	TCGGAGTGTC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
	701	CTGCCCTCGT	TCAGTTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATGTT
	751	CGTCAATCGT	TGTGTTCTCG	A		

45	1	ISYWASSSLD	FLEVDTAPLI	FLPLLPKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	FLTFXTSCPP	RSNRPYQYRR
	101	LRLYAFHAPL	ITEFFVGFAP	XVDARNVYAQ	IGGVGVTHLR	RNRREFGLFL
	151	NHGRIDIDRL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGEM	AADIAQTCRT
50	201	EQRVGNVQVQ	RIGIGVSEQP	FFKWFDFNSAK	YQLSAFGQLV	DIVALSDTDV
	251	RHRLCS*				

		10	20	30	40	50	60
	orf122a.pep	ISYWASSSLDFLEVD	TAPLIFLPLLPKAS	MKKLMVEFVPM	MPMYSFSGTN	STAFSAAMRLS	
55	orf122-1	ISYWASSSPDFLEVD	TAPLIFLPLLPKAS	MKKLMVEFVPM	PIYSFSGTN	STAFSAAMRLS	
		10	20	30	40	50	60
		70	80	90	100	110	120
60	orf122a.pep	SSCVVIFLSFGKPY	QQTAAILTF	FXTS	SCPPSRNPY	QQYRRLRL	YAFHAP
	orf122-1	SSCVVIFLSFGKPY	QQTAAILTF	FCTSCPPSRN	AYQQYRRLRL	YAFHPPEIA	EEFVGFAF

-441-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf122a.pep	XVDARNVYAQIGGDVGTHLRNMREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
	orf122-1	DVDARNVYAQIGGDVGTHLRNVREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
10	orf122a.pep	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWDNFNSAKYQLSAFGQLV					
	orf122-1	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWDNFNSAKYQLSAFGQLV					
		190	200	210	220	230	240
15	orf122a.pep	DIVALSDTDVRHRLCSX					
	orf122-1	DIVALSDTDVRHRLCSX					
		250					
20							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from *N.gonorrhoeae*:

25	orf122.pep	TAFSAALRLSPSXLVIFLSFGKPYQQTAAI	30
	orf122ng	FLPLLKASMKKLMVEPVMPMYSFSGTNSAFSAAMRLSSSCVVIFLSFGKPYQQTAAI	80
	orf122.pep	LTFEFTSCPPRSNAYQQYRRLRLRYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR	90
30	orf122ng	LTFEFTSWPPRSNPYQQYRRLRLRYAFHPPEIAEFFVGFAFDIDARNIDTQIGGDVGTHLR	140
	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT	150
35	orf122ng	NVRCEFGFLCNHGRIDIDHPLTLRLNALIRRTQKDAAVRIFELCGGVGKMAADVAQTCRT	200
	orf122.pep	EQRVGNVQQRIGIGVSEQPFFKWDNFNSAKYQ	182
	orf122ng	EQRVGNVQQRVGIRMPEQPFFKWDNFNSAKYQLSAFGQLVDIVALSDTDIRHRLCS	256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

40	1	ATGTCGTACC	GGGCAAGCAG	TTCGCCGAT	TTTTTGGAGG	TTGAAACCGC
	51	GCCTTTGATT	TTTTTACCGC	TTTGGCCCAA	GGCTTCGATG	AAGAAATTGa
	101	tgGTCGAACC	GgtaCCGATG	CCGATGTATT	CGTTTCGGG	TACGAATCG
	151	ACTGCTTTTT	CGGCGGCGAT	GCGCttgAgt	TCgtcttgcg	TcgTCATATT
45	201	TTTAtccttt	gGGAaccct	atcaAcaAAc	agccgccatC	TTAACATTTT
	251	TTTGCACGtc	ctggccgcgcg	cgttcaAATc	cgtaccaGca	ataccgcgcg
	301	ctgcgcctCT	AtgcCTTCCA	TCCGCCCGAG	ATAGCCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GATattGACG	CACGAAATAT	CGatacCCaA	atcggcgcGCG
	401	ATGTTGGCAC	GCATTTGCGG	AATGTGCGGT	GCGAGTTTGG	GTTTCTGTGC
50	451	AATCACGGTC	GTATCGACAT	TGACCACCTG	CCAACCCTGC	GCCTGAACGC
	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCATC	TTTGAACCTCT
	551	GCGGCGGTGT	CGGAAAAATG	GCTGCCGATG	TCGCCCAAAC	CTGCCGCACC
	601	GAGCAGCgcg	tcggtaaCGG	CGTGACGAG	cgcgTcgGCA	TCCGAATGCC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
55	701	CTGCCTTCGG	TCAATTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATATT
	751	CGTCATCGTT	TGTGTTCTG	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

	1	MSYRASSSPD	FLEVETAPLI	FLPLLKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFEFTSWPP	RSNPYQQYRR
60	101	LRLYAFHPPE	IAEFFVGFAF	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC
	151	NHGRIDIDL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCRT
	201	EQRVGNVQQR	RVGIRMPEQP	FFKWDNFNSAK	YQLSAFGQLV	DIVALSDTDI
	251	RHRLCS*				

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122-1.pep	ISYWASSSPDFLEVD	TAPLI	IFLPLLPKAS	MKKLMVE	PVPMPIYS	SFSGTNSTAFSAAMRLS
5	orf122ng	MSYRASSSPDFLEV	ETAPLI	IFLPLLPKAS	MKKLMVE	PVPMPIYS	SFSGTNSTAFSAAMRLS
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf122-1.pep	SSCVVIFLSFGKPY	QQTAA	ILTFECT	SCPPRSN	AYQQYRRL	RLYAFHPPEIAEFFVGFAF
	orf122ng	SSCVVIFLSFGKPY	QQTAA	ILTFECT	SWPPRSN	PYQQYRRL	RLYAFHPPEIAEFFVGFAF
		70	80	90	100	110	120
		130	140	150	160	170	180
15	orf122-1.pep	DVDARNVYAQIGG	DVGTHLR	NVRREFG	FLCNHGR	IDIDRLPT	LRLNALIRRTQKDAAVRI
	orf122ng	DIDARNIDTQIGG	DVGTHLR	NVRCEFG	FLCNHGR	IDIDHLP	TLRLNALIRRTQKDAAVRI
		130	140	150	160	170	180
20		190	200	210	220	230	240
	orf122-1.pep	FELCGGVGEMAAD	IAQTCR	TEQRVGN	GVQQRIG	IGVSEQP	FFKWFDFNSAKYQLSAFGQLV
	orf122ng	FELCGGVGKMAAD	VAQTCR	TEQRVGN	GVQQRV	GIRMPEQ	FFKWFDFNSAKYQLSAFGQLV
25		190	200	210	220	230	240
		250					
	orf122-1.pep	DIVALSDTDVRH	RRLCSX				
30	orf122ng	DIVALSDTDIRH	RRLCSX				
		250					

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 95

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

35	1	..GCCGGCGCGA	GTGCGAACAA	CATTTCGCG	CGTTTTGCGG	AAACACCCGT
	51	CGCTGTCAGC	GTTACCCTGA	TCGGCACGGT	ACTTGCCGTC	ATGCTGCCCG
	101	TTACCGAATA	TGAAACTTC	CTGCTGCTTA	TCGGCTCGGT	ATTTGCGCCG
	151	ATGGGGCGGA	TTTGATTGC	CGACTTTTTC	GTCTTGAAAC	GGCGTGA

This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

40	1	..AGASANNISA	RFAETPVAVS	VTLIGTVLAV	MLPVTEYENF	LLIGSVFAP
	51	MGGFDCRLFR	LETA*			

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

	1	ATGTCGGGCA	ATGCCTCCTC	TCCTTCATCT	TCCTCCGCCA	TCGGGCTGAT
	51	TTGGTTCCGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCACG	GGTACGCTGC
45	101	TTGCGCCTTT	GGGCTGGCAG	CGCGGTCTGG	CGGCTCTACT	TTTGGGTCAT
	151	GCCGTCGGCG	GCGCGCTGTT	TTTTGCGGCG	GCGTATATCG	GCGCACTGAC
	201	CGGACGCAGC	TCGATGGA	AAACGCGGTT		
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCCCG	CTGGACGGCG
	301	GTGATGATTT	ACGCCGGCGC	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
50	351	GTGGGACGGC	GAATCTTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCGCTGA
	401	TTGTGCTGTG	GCTGGTTTTT	GGCGCACGCA	AAACAGGCGG	GCTGAAAACC
	451	GTTTCGATGC	TGCTGATGCT	GTGGCGGTT	CTGTGGCTGA	GTGCCGAAGT
	501	CTTTTCCACG	GCAGGCAGCA	CCGCCGCACA	GGTTTCAGAC	GGCATGAGTT
	551	TCGGAACGGC	AGTCGAGCTG	TCCGCCGTGA	TGCCGCTTTC	CTGGCTGCCG
55	601	CTTGCCGCGG	ACTACACGCG	CCACGCGCGC	CGCCCGTTTG	CGGCAACCCCT
	651	GACGGCAACG	CTCGCCTACA	CGCTGACCGG	CTGCTGGATG	TATGCCTTGG
	701	GTTTGGCAGC	GGCGTTGTTT	ACCGAGAGAA	CCGACGTGGC	AAAAATCCTG
	751	CTGGGCGCAG	GTTTGGGTGC	GCGAGGCATT	TTGGCGGTGC	TCCTCTCCAC

5  
 801 CGTTACCACA ACGTTTCTCG ATGCCTATTC CGCCGGCGCG AGTCCGAACA  
 851 ACATTTCCGC GCGTTTTCGC GAAACACCCG TCGCTGTCGG CGTTACCCCTG  
 901 ATCGGCACGG TACTTGCCGT CATGCTGCCC GTTACCGAAT ATGAAAACCT  
 951 CCTGCTGCTT ATCGGCTCGG TATTGCGGCC GATGGCGGCG GTTTTGATTG  
 1001 CCGACTTTT CGTCTTGAAA CGGCGTGAGG AGATTGAAGG CTTTGACTTT  
 1051 GCCGGACTGG TTCTGTGGCT TCGGGGCTTC ATCCTCTACC GCTTCCTGCT  
 1101 CTCGTCCGGC TGGGAAAGCA GCATCGGTCT GACCGCCCC GTAATGTCTG  
 1151 CCGTTGCCAT TGCCACCGTA TCGGTACGCC TTTTCTTTAA AAAAACCCAA  
 1201 TCTTTACAAA GGAACCCGTC ATGA

10 This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

15  
 1 MSGNASSPSS SSAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH  
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KRGSVLFSVA NMLQLAGWTA  
 101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT  
 151 VSMLLMLLAV LWLSAEVFST AGSTAAQVSD GMSFGTAVEL SAVMPLSWLP  
 201 LAADYTRHAR RPFATLTAT LAYTLTGCMW YALGLAAALF TGETDVAKIL  
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGA SANNISARFA ETPVAVGVTL  
 301 IGTVLAVMLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEGFDF  
 351 AGLVLWLAGF ILYRFLSSG WESSIGLTAP VMSAVAIATV SVRLFFKKTQ  
 401 SLQRNPS\*

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N. meningitidis*:

25  
 orf125.pep AGASANNISARFAETPVAVSVTLLIGTVLAV  
 orf125a KILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAVVGTLLAV  
 250 260 270 280 290 300  
 30  
 orf125.pep MLPVTEYENFLLIGSVFAPMGGFDCRLFRLETAX  
 orf125a LLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEG  
 310 320 330 340

35 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

40  
 1 ATGTCGGGCA ATGCCTCCTC TCNTTCATCT TCCGCGGCCA TCGGGCTGAT  
 51 TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACACTGC  
 101 TTGCGCCTTT GGGCTGGCAG CGCGGTCTGG CNGCTCTGCT TTTGGGTCAT  
 151 CCGTCTGGCG GCGGCTGTT TTTTGGCGCG GCGTATATCG GCGCACTGAC  
 201 CGGACNCANC TCGATGGAAA GCGTGGCCTT GTCGTTCCGC AAACGCGGTT  
 251 CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGCGCGG CTGGACGGCG  
 301 GTGATGATT ACGCCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT  
 351 GTGGGACGGC GAATCTTTTG TCTGGTGGGC ATTGGCAAAC GGCGCGCTGA  
 401 TTGTGCTGTG GCTGGTTTTC GGCGCACGCA AAACAGCGCG GCTGAAAACC  
 451 GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGGCTGA GTGCCGAANT  
 501 NTTTTCCACG CGAGGCAGCA CCGCCGCANN GGTNNCAGAC GGCATGAGTT  
 551 TCGGAACGGC AGTCGAGCTG TCCGCGGTNA TGCCGCTTTC TTGGCTGCCG  
 601 CTGGCCGCCG ACTACACGCG CCACGCGCGC CGCCCGTTTG CGGCAACCC  
 651 GACGGCAACG CTCGCCTACA CGCTGACCGG CTGCTGGATG TATGCCTTGG  
 701 GTTTGGCAGC GCGGTTGTT ACCGGAGAAA CCGACGTGGC AAAAATCCTG  
 751 CTGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGTGC TCCTGTGCGC  
 801 CGTTACCACC ACTTTTCTCG ATGCNTACTC CGCCGGCGTA AGTGCCAACA  
 851 ATATTTCCGC CAAACTTTTC GAAATACCNA TCGCCGTTGC CGTCGCCGTT  
 901 GTCGGCACAC TGCTTGCCGT CCTCCTGCCC GTTACCGAAT ATGAAAACCT  
 951 CCTGCTGCTT ATCGGCTCGG TATTGCGGCC GATGGCGGCG GTTTTGATTG  
 1001 CCGACTTTT CGTCTTGAAG CCGCGTGAGG AGATTGAAGG C..

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

1 MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH  
 51 AVGGALFFAA AYIGALTGX SMESVRLSFG KRGSVLFSVA NMLQLAGWTA

-444-

101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT  
 151 VSMLLMLLAV LWLSAEXFST AGSTAAXVXD GMSFGTAVEL SAVMPLSWLP  
 201 LAADYTRHAR RPPAATLTAT LAYTLTGCM YALGLAAALF TGETDVAKIL  
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGV SANNISAKLS EIPIAVAVAV  
 5 301 VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
10	orf125a.pep	MSGNASSXSSSA	IGLIWFGAAVS	IAEISTGTLLAP	LGWQRLAALLL	GHAVGGALFFAA	
	orf125-1	MSGNASSPSSSA	IGLIWFGAAVS	IAEISTGTLLAP	LGWQRLAALLL	GHAVGGALFFAA	
		10	20	30	40	50	60
15	orf125a.pep	70	80	90	100	110	120
	orf125-1	70	80	90	100	110	120
20	orf125a.pep	130	140	150	160	170	180
	orf125-1	130	140	150	160	170	180
25	orf125a.pep	190	200	210	220	230	240
	orf125-1	190	200	210	220	230	240
30	orf125a.pep	250	260	270	280	290	300
	orf125-1	250	260	270	280	290	300
35	orf125a.pep	310	320	330	340		
	orf125-1	310	320	330	340	350	360

#### Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from *N.gonorrhoeae*:

	orf125.pep	AGASANNISARFAETPVAVS	TLIGTVLAV	30
	orf125ng	KILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVTLIRT	VLAV	308
50	orf125.pep	MLPVTEYENFLLLIGSVFAPM-GGFD	CRLFRLETA	64
	orf125ng	MLPVTEYKNFLLIRSVFGPMAGGFD	CRLFCLKTA	343

An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

1 MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH  
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA  
 101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT  
 151 VSMLLMLLAV LWLSVEFEAS SGTNAAPAVS DGMTFGTAVE LSAMPLSWL  
 201 PLAADYTRQA RPPAATLTAT TLAYTLTGCM MYALGLAAAL FTGETDVAKI  
 251 LLGAGLGITG ILAVVLSTVT TFLDTYSAG ASANNISARF AEIPVAVGV  
 301 LIRTVLAVML PVTEYKNFLL LIRSVFGPMA GGFDCLFCL KTA\*



Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

```

1  ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCGCCGCCA TCGGGCTGGT
51  TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
101 TCGCCCCCTT GGGCTGGCAG CGCGGCTGG CGGCCCTGCT TTTGGGTCAT
151 GCGCTCGGCG GCGCGCTGTT TTTTGGCGCG GCGTATATCG GCGCACTGAC
201 CGGACGCAGC TCGATGGAAA GTGTGCGCCT GTCGTCGGC AAATGCGGTT
251 CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
301 GTGATGATTT ACGTCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT
351 GTGGGACGGC GAATCCTTTG TCTGGTGGG ATTGGCAAAC GCGCACTGA
401 TCGTGCTGTG GCTGTTTTC GCGGCACGCA GAACGGGCGG GCTGAAAACC
451 GTTTCGATGC TGCTGATGCT GCTTGCCGTG TTGTGGTTGA GCGTCGAAGT
501 GTTCGCTTCG TCCGGCACA ACGCCGCGCC CGCGGTTTCA GACGGCATGA
551 CCTTCGGAAC GGCAGTCGAA CTGTCCGCG TCATGCCGCT TTCCTGGCTG
601 CCGCTGGCCG CCGACTACAC GCGCCAAGCA CGCGCCCGCT TTGCGGCAAC
651 CCGTACGGCA ACGCTCGCCT ATACGCTGAC GGGCTGCTGG ATGTATGCCT
701 TGGGTTTGGC GCGGCTCTG TTTACCGGAG AAACCGACGT GGCGAAAATC
751 CTGTGGGCG CGGGCTTGG CATAACGGG ATTCTGGCAG TCGTCTCTC
801 CACCGTTACC ACAACGTTTC TCGATACCTA TTCCGCCGGC GCGAGTGC GA
851 ACAACATTT CCGCGGTTTT GCGGAAATAC CCGTCGCTGT CCGCGTTACC
901 CTGATCGGCA CGGTGCTTGC CGTCATGCTG CCGGTTACCG AATATAAAAA
951 CTTCTGCTG CTTATCGGCT CGGTATTTGC GCCGATGGCG GCGGTTTTGA
1001 TTGCCGACTT TTTCTCTTA AAACGGCGTG AGGAGATTGA AGGCTTTGAC
1051 TTTGCCGGAC TGTTCTGTG GCTGGCAGG TTCATCCTCT ACCGCTTCTT
1101 GCTCTCGTCC GGTTCGGAAA GCAGCATCGG TCTGACCGCC CCCGTAATGT
1151 CTGCCGTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAACC
1201 CAATCTTTAC AAAGGAACCC GTCATGA

```

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

```

1  MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLWQ RGLAALLLGH
51  AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
151 VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
201 PLAADYTRQA RRPFAATLTA TLAYTLTGCW MYALGLAAAL FTGETDVAKI
251 LLGAGLGITG ILAVVLSTVT TTFLDTYSAG ASANNISARF AEIPVAVGVT
301 LIGTVLAVML PVTEYKNFL LIGSVFAPMA AVLIADFFVL KRREEIEGFD
351 FAGLVWLWAG FIFYRFLSS GWESSIGLTA PVMSAVAIAT VSVRLFFKKT
401 QSLQRNPS*

```

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

```

40  orf125-1.pep 10 20 30 40 50 60
MSGNASSPSSSSAIGLIWFGAAVSVIAEISTGTLLAPLWQ RGLAALLLGHAVGGALFFAA
|||||:||||:|||||
orf125ng-1 MSGNASSPSSSSAIGLVWFGAAVSVIAEISTGTLLAPLWQ RGLAALLLGHAVGGALFFAA
10 20 30 40 50 60

45  orf125-1.pep 70 80 90 100 110 120
AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
|||||:||||:|||||
orf125ng-1 AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYVGATVSSALGKVLWDG
70 80 90 100 110 120

50  orf125-1.pep 130 140 150 160 170 179
ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFSTAGSTAAQ-VS
|||||:||||:|||||
orf125ng-1 ESFVWWALANGALIVLWLVFGARRTGGLKTVSMLLMLLAVLWLSVEVFASSGTNAAPAVS
130 140 150 160 170 180

55  orf125-1.pep 180 190 200 210 220 230 239
DGMSEGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCWYALGLAAAL
|||:|||||:|||||
orf125ng-1 DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTGCWYALGLAAAL
190 200 210 220 230 240

60  orf125-1.pep 240 250 260 270 280 290 299
FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVT
|||||:||||:|||||
orf125ng-1 FTGETDVAKILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVT
250 260 270 280 290 300

```

		300	310	320	330	340	350	359
	orf125-1.pep	LIGTVLAVMLPVTEYENFLL	LIGSVFAPMAAVLIADFFVLKRREEIEG	FDFAGLVLWLAG				
5	orf125ng-1	LIGTVLAVMLPVTEYKFN	LIGSVFAPMAAVLIADFFVLKRREEIEG	FDFAGLVLWLAG				
		310	320	330	340	350	360	
		360	370	380	390	400		
	orf125-1.pep	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX						
10	orf125ng-1	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX						
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
20	51	GTTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TAGCCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	A.ACGGTCGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
25	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGT.ACGBA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTAA	GACGGCATCT	ACCTGCCGAC	CGAAGC.CAG
	451	CTCGACGGGC	GGCAATTATA	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GCCTGCAAG..

This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

1	MTRIAILGGG	LSGRILTALQL	AEQGYQIALF	DKSCRRGEHA	AAYVAAAMLA
51	PAAXTVEATP	EVVRLGRQSI	PLWRGIRCRL	NHTMMQENG	SLIVWHGQDK
101	PLSSEFVRHL	KRGXTDDEI	VRWRADDIAE	REPQLGGRFX	DGIYLPTEXQ
151	LDGRQLXSAL	ADALDELNVP	CHWEHECVPE	ACK...	

Further work revealed the complete nucleotide sequence <SEQ ID 811>:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TAGCCGCCGC	CATGCTCGCG
40	151	CCTGCAGCGG	AAGCGGTCGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TGTGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGTAGCGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
45	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCCGAC	CGAAGGCCAG
	451	CTCGACGGGC	GGCAAATATT	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GGCCTGCAAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCCGCG	GCTACGGCGC	AAAAACCGCG
	601	TGGAACCAAT	CCCCGAGCA	CACCAGCACC	CTGCGCGGCA	TACGCGGCGA
50	651	AGTGGCGCGG	GTTTACACAC	CCGAAATCAC	GCTCAACCGC	CCCGTGCCTC
	701	TGCTCCATCC	GCGTTATCCG	CTCTACATCG	CCCCGAAAGA	AAACCACGTC
	751	TTCGTCATCG	GCGCGATCCCA	AATCGAAAGC	GAAAGCCAAG	CCCCCGCCAG
	801	CGTGCGTTCA	GGGTTGGAAC	TCTTGTCCGC	ACTCTATGCC	ATCCACCCCG
	851	CCTTCGGCGA	AGCCGACATC	CTCGAAATCG	CCACCGGCTT	GCGCCCCACG
55	901	CTCAACCACC	ACAACCCCGA	AATCCGTTAC	AACCGCGCCC	GACGCCTGAT
	951	TGAAATCAAC	GGCCTTTTCC	GCCACGGTTT	CATGATCTCC	CCGCGCTTAA
	1001	CCGCGCGCGC	CGCCAGATTG	GCAGTGGCAC	TGTTTGACGG	AAAAGACGCG
	1051	CCCGAACGCG	ATAAAGAAAG	CGGTTTGGCG	TATATCCGAA	GACAAGATTA
	1101	A				

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51  PAEAVEATP EVVRLGRQSI PLWRGIRCL NTHMTMQENG SLIVWHGQDK
101 PLSSEFVRHL KRGGVADDEI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
5  151 LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAQYDWLI DCRGYAKTA
201 WNQSPEHTST LRGIERGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV
251 FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
301 LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAARL AVALFDGKDA
351 PERDKESGLA YIRRQD*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N. meningitidis*:

```

15  orf126.pep      10      20      30      40      50      60
      |||||
orf126a      MTRIAILGGG LSGRLTALQLAEQGYQIALFDKGCRRGEHAAYVAAAMLAPAEAVEATP
      10      20      30      40      50      60

20  orf126.pep      70      80      90      100     110     120
      |||||
orf126a      EVVRLGRQSIPLWRGIRCLNTHMTMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI
      70      80      90      100     110     120

25  orf126.pep      130     140     150     160     170     180
      |||||
orf126a      VRWRADDIAEREPLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE
      130     140     150     160     170     180

```

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCNGGAA GGCTGACCGC
51  ACTGCAGCTT GCAGAACAAAG GTTATCAGAT TGCACTTTTC GATAAAGGCT
101 GCCGCCGGGG CGAACACGCC GCCGCCTATG TTGCCGCCGC CATGCTCGCG
35  151 CCTGCGGCGG AAGCGGTCGA AGCCACGCCT GAAGTGGTCA GGCTGGGCAG
201 CCAGANCATC CCGCTTTGGC GCGGCATCCG ATGCCATCTG AAAACGCCTG
251 CCATGATGCA NGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAA
301 CCTTTATCCA ACGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
40  351 TGACNAAATC GTCCGTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
401 AACTCGGCGG ACGTTTTTC ACGGCATCT ACCTGCCGAC CGAAGGCCAG
451 CTCGACGGGC GGCAAATATT GTCTGCACTT GCCGACGCTT TGGACGAAT
501 GAACGTCCCC TGCCATTGGG AACACGAATG TGCCCCGAA GACTTGCAAG
551 CCCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAACCGCG
601 TGGAAACCAAT CCCCCGANNA NACCAGCACC CTGCGCGGCA TACGCGGCGA
45  651 AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
701 TGCTACACCC GCGCTATCCG CTNTACATCG CCCCAGAAAG AAACNCGTC
751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CACCTGCCAG
801 CGTGCCTTCC GGGCTGGAAC TCTTATCCGC ACTCTATGCC GTCCACCCCG
851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCACCGGCCT GCGCCCCACG
50  901 CCTCAATCACC ACAACCCGGA AATCCGTAC AACCGCGCCC GACGCTGAT
951 TGAAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCCGCCGTAA
1001 CCGCGGCCGC CGTCAGATTG GCAGTGGCAC TGTGTGACGG AAAAGANGCG
1051 CCCGAACGCG ATGAAGAAAG CGGTTTGCGC TATATCCGAA GACAAGATTA
1101 A

```

This encodes a protein having amino acid sequence <SEQ ID 814>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51  PAEAVEATP EVVRLGRQXI PLWRGIRCL KTPAMMXENG SLIVWHGQDK
101 PLSNEFVRHL KRGGVADDXI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECAPE DLQAQYDWLI DCRGYAKTA
60  201 WNQSPXXTST LRGIERGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENXV

```

251 FVIGATQIES ESQAPASVRS GLELLSALYA VHFAFGEADI LEIATGLRPT  
 301 LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKXA  
 351 PERDEESGLA YIRRQD\*

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

5	orf126a.pep	10	20	30	40	50	60
	orf126-1	10	20	30	40	50	60
10	orf126a.pep	70	80	90	100	110	120
	orf126-1	70	80	90	100	110	120
15	orf126a.pep	130	140	150	160	170	180
	orf126-1	130	140	150	160	170	180
20	orf126a.pep	190	200	210	220	230	240
	orf126-1	190	200	210	220	230	240
25	orf126a.pep	250	260	270	280	290	300
	orf126-1	250	260	270	280	290	300
30	orf126a.pep	310	320	330	340	350	360
	orf126-1	310	320	330	340	350	360
35	orf126a.pep	310	320	330	340	350	360
	orf126-1	310	320	330	340	350	360
40	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					
45	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

50	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHAAAYVAAAMLAPAAEAVEATP	60
55	orf126.pep	EVVRLGRQSIPLWRGIRCLNTHMTMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI	120
	orf126ng	EVIRLGRQSIPLWRGIRCLNTHMTMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI	120
	orf126.pep	VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE	180
	orf126ng	VRWRADDEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQ	180

60 An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTROGEHA AAYVAAAMLA

51 PAEEAVEATP EVIRLGRQSI PLWRGIRCRL NTLTMMQENG SLIVWHGQDK  
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ  
 151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA  
 201 WNQSPEHTST LRGIRGEVRG FTRPKSRSTA PCACCTRAIR STSPRKKTTTS  
 251 SSSARPKSKA KAKPPPAYVP GWNSYPRSMPT STPPSAKPTS SKWRPGLRPT  
 301 LNHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA  
 351 PERDEESGLA YIGRQD\*

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

1 ATGACCCGTA TCGCCGTCCT CGGAGGCGGC CTTTCCGGAA GGCTGACCGC  
 51 ATTGCAGCTT GCAGAACAAAG GTTATCAGAT TGAACCTTTC GACAAGGGCA  
 101 CCCGCCAAGG CGAACACGCC GCCGCCTATG TTGCCGCCGC GATGCTCGCG  
 151 CCTGCCGCGG AAGCGGTCTGA GGCAACGCCG GAAGTCATCA GGCTGGGCAG  
 201 GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCTCA  
 251 CGATGATGCA GGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG  
 301 CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA  
 351 TGACGAAATC GTCGTTGGC GCGCCGATGA AATCGCCGAA CCGGAACCGC  
 401 AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG  
 451 CTCGACGGGC GGCAAAATAT GTCTGCACTT GCCGACGCTT TGGACGAAC  
 501 GAACGTCCTT TGCCATTGGG AACACGAATG CGCCCCCAA GACCTGCAAG  
 551 CCCAATACGA CTGGGTAATC GACTGCCGGG GCTACGGCGC GAAAACCGCG  
 601 TGAACCAAT CCCCGAGCA CACCAGCACC TTGCGCGGCA TACGCGGCGA  
 651 AGTGGCGCGG GTTACACGC CCGAAATCAC GCTCAACCGC CCGTGCGCC  
 701 TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCAGAAAG AAACCACGTC  
 751 TTCGTATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCCGCCAG  
 801 CGTACGTTCC GGGCTGGAAC TCTTATCCGC GCTCTATGCC GTCCACCCCG  
 851 CTTTCGGCGA AGCCGACATC CTCGAAATCG CCGCCGGCCT GCGCCCCACG  
 901 CTCAACCACC ACAACCCCGA AATCCGCTAC AGCCGCGAAC GCCGCCTCAT  
 951 CGAAATCAAC GGCTTTTCC GGCACGGCTT TATGATTTC CCGCCGTAA  
 1001 CCGCCGCGCG CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG  
 1051 CCCGAACGTG ATGAAGAAAG CGGTTTGCG TATATCGGA GACAAGATTA  
 1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTRQGEHA AAYVAAAMLA  
 51 PAEEAVEATP EVIRLGRQSI PLWRGIRCRL NTLTMMQENG SLIVWHGQDK  
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ  
 151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA  
 201 WNQSPEHTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV  
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT  
 301 LNHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA  
 351 PERDEESGLA YIGRQD\*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

10 20 30 40 50 60  
 orf126-1.pep MTRIAVLGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEEAVEATP  
 45 orf126ng-1 MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHAAAYVAAAMLAPAAEEAVEATP  
 10 20 30 40 50 60  
 70 80 90 100 110 120  
 orf126-1.pep EVVRLGRQSIPLWRGIRCRLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI  
 50 orf126ng-1 EVIRLGRQSIPLWRGIRCRLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI  
 70 80 90 100 110 120  
 130 140 150 160 170 180  
 orf126-1.pep VRWRADDIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE  
 55 orf126ng-1 VRWRADEIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQ  
 130 140 150 160 170 180  
 190 200 210 220 230 240  
 orf126-1.pep GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP  
 60 orf126ng-1 DLQAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHPRYP  
 190 200 210 220 230 240  
 65

-450-

		250	260	270	280	290	300
	orf126-1.pep	LYIAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAIHPAFGEADILEIATGLRPT					
5	orf126ng-1	LYIAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAVHPAFGEADILEIAAGLRPT					
		250	260	270	280	290	300
	orf126-1.pep	LNHHNPEIRYNRRRLIEINGLFRHGFMI	310	320	330	340	350
10	orf126ng-1	LNHHNPEIRYSRERRLIEINGLFRHGFMI	310	320	330	340	350
		310	320	330	340	350	360
15	orf126-1.pep	YIRRQDX					
	orf126ng-1	YIGRQDX					

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

	gi 2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
20	Length = 327
	Score = 169 bits (423), Expect = 3e-41
	Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)
	Query: 3 RIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHXXXXXXXXXXXXXXXXXXXXX 62
25	Sbjct: 2 RILVNGACVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGCMLAPWCERESAEPEV 60
	Query: 63 IRLGRQSIPLWRGIRCLNLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEIVR 122
30	Sbjct: 61 LTLGRLAADWWEAA-----LPGHVHRRGTLVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113
	Query: 123 WRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQDL 182
35	Sbjct: 114 -----IAALEPDLAGRFRRALFFRQEAHLDPQALAAALAGLEDARMRLTLG---VVGES 165
	Query: 183 QAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPEITLNRPVRLHLHPRYPLY 242
40	Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVRGEMLCVETTEVSLSRPVRLHLHPRHPIY 218
	Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAVHPAFGEADILEIAAGLRPTLN 302
45	Sbjct: 219 IVPRDKNRFMVGATMIESDDGGPITARSLMELNAAAYAMHPAFGEARVTETGAGVRPAYP 278
	Query: 303 HHNPEIRYSRERRLIEINGLFRHGFMI 331
	Sbjct: 279 DNLP--RVTQEGRTLHVNGLYRHGFLLAP 305

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

50	819>:
	1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
	51 GATATTGCTT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
	101 TTGAGAAAGC AAAGATAAAT GCAAGTCCGGG CAGCCTTGTT AGAAAATGCA
55	151 CATTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGGTTTA AACAAACATC
	201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
	251 GTTTGAATGG AATCGTCGCG CGGG..GCTT TAGACAGTAA ATTCATGTTG
	301 AAGGCGGTAG CCATAGATAA AGATAAAAAT CCTTTTATTA TTAAGATGAA
	351 TGAATCTA GTAACCTTTA aTTTGAAGA AGTCCGCCAG TTCGTGTAGT
60	401 GACGGGCTGG ATTATTTTAA AGGAAATGAT AAGGACTGCA AGTTACTTAA
	451 GTAG

Further work revealed the following DNA sequence <SEQ ID 821>:

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N*.

*meningitidis:*

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

This encodes a protein having amino acid sequence <SEQ ID 824>:

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLEN  
51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKEMLK  
101 AVAIDKDKNP FIIKMNNELV TFCCKKSASS CSDGLDYFKG NDKDCKLIK\*

-452-

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

```

      10      20      30      40      50      60
orfl27a.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINTVRAALLEN AHFMEKFY LQN
5 orfl27-1   MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFY LQN
      10      20      30      40      50      60

      70      80      90     100     110     120
orfl27a.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
10 orfl27-1   GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
      70      80      90     100     110     120

      130     140     150
orfl27a.pep TFICKKSASSCS DGLDYFKGNDKDKCKLLKX
15 orfl27-1   TFICKKSASSCS DGLDYFKGNDKDKCKLLKX
      130     140     150

```

## 20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from *N.gonorrhoeae*:

```

orfl27.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFY LQN 60
25 orfl27ng MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAAFLENA HFMEKFY LQN 60

orfl27.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGI VARXALDSKFMLKAVAIDKDKNPFI IKMNENL 120
30 orfl27ng GRFKQTSTKWPSLPIKEAEGFCIRLNGI -ARGALDSKFMLKAVAIDKDKNPFI IKMNENL 119

orfl27.pep VTFICKKSASSCS DGLDYFKGNDKDKCKLLK 150
orfl27ng VTFICKKSASSCS DRLDYFKGNDKDKCKLLK 149

```

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

```

35 1 ATGACTGATA ATCGGGGGTT TACACTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGATTTA AACAAACATC
201 TACCAAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
40 251 GTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 826>:

```

45 1 MTDNRGFTLV ELISVVLILSVLALIVYPSY RNYVEKAKIN AVRAAFLENA
51 HMEKFY LQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDKNP FIIKMENLV TFICKKSASS CSDRLDYFKG NDKDKCKLLK*

```

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

```

      10      20      30      40      50      60
50 orfl27-1.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFY LQN
orfl27ng-1 MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLEN AHFMEKFY LQN
      10      20      30      40      50      60

      70      80      90     100     110     120
55 orfl27-1.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
orfl27ng-1 GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFI IKMNENLV
      70      80      90     100     110     120
60

```



-453-

		130	140	150
orf127-1.pep	TFICKKSASSCS	DGLDYFKGNDKDCKLLKX		
orf127ng-1	TFICKKSASSCS	DGLDYFKGNDKDCKLLKX		
		130	140	150

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 10 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

	1	..GTGTCGCTGG	CTTCGGTGAT	TGCCTCTCAA	ATCTTCCTTT	ACGAAGATTT
	51	CAACCAAATG	CGGAAAACCG	GTGGAGCTAT	CTGCGGTTTT	CTTGTCCAAT
15	101	ATTTATCTGG	GGTTTCAGCA	GGGGTATTTT	GATTTGAGTG	CCGACGAGAA
	151	CCCCGACTG	CATATCTGGT	CTTTGGCAGT	AGAGGAACAG	TATTACCTCC
	201	TGTATCCCCT	TTTGCTGATA	TTTGCTGCA	AAAAAACAA	ATCGCTACGG
	251	GTGCTGCGTA	ACATCAGCAT	CATCCTGTTT	TTGATTTTGA	CTGCCTCATC
	301	GTTTTTGCCA	AGCGGGTTTT	ATACCGACAT	CCTCAACCAA	CCCAATACTT
	351	ATTACCTTTC	GACACTGAGG	TTTCCCGAGC	TGTTGGCAGG	TTCGCTGCTG
20	401	GCGGTTTACG	GGCAAACGCA	AAACGGCAGA	CGGCAAACAG	CAAATGGAAA
	451	ACGGCAGTTG	CTTTCATCAC	TCTGCTTCGG	CGCATTGCTT	GCCTGCCTGT
	501	TCGTGATTGA	CAAACACAAT	CCGTTTATCC	CGGGAATGAC	CCTGCTCCTT
	551	CCCTGCCTGC	TGACGGCACT	GCTTATCCGG	AGTATGCAAT	ACGGGACACT
	601	TCCGACCCGC	ATCCTGTCCG	CAAGCCCCAT	CGTATTTGTC	GGCAAAATCT
25	651	CTTATTCCTT	ATACCTGTAC	CATTGGATT	TTATGCTTT	CGCTCCGCTC
	701	ATTAGAGGCG	GGAAACAGCT	CGGACTGCCT	GCCG..	

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

	1	..VSLASVIASQ	IFLYEDFNQM	RKTVELSAVF	LSNIYLGFOQ	GYFDLSADEN
	51	PVLHIWSLAV	EEQYYLLYPL	LLIFCCKKTK	SLRVLRNISI	ILFLILTASS
30	101	FLPSGFYTDI	LNQPNYYLYS	TLRFPELLAG	SLLAVYGTQ	NGRRQTANGK
	151	RQLSSSLCFG	ALLACLFVID	KHNPFIPIGT	LLLPCLLTAL	LIRSMQYGT
	201	PTRILSASPI	VFVGKISYSL	YLYHWIFIAF	APLIRGGKQL	GLPA..

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

	1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACGGATTGC	GGGCCGTCGC
35	51	CGTGCTATCC	GTGATGATTT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
	101	GATTCCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC
	151	GGCATCATTC	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTTT	TCCGGGATTT
	201	TTATACCCGC	AGGATTAAGC	GGATTATATC	TGCCTTTATT	GCGGCCGTGT
40	251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
	301	CAAATGCGGA	AAACCGTGGA	GCTTCTGCG	GTTTTCTTGT	CCAATATTTA
	351	TCTGGGGTTT	CAGCAGGGGT	ATTTGATTTT	GAGTGCCGAC	GAGAACCCCG
	401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCCTGTAT
	451	CCCTTTTTC	TGATATTTTG	CTGCAAAAAA	ACCAAATCGC	TACGGGTGCT
	501	GCGTAACATC	AGCATCATCC	TGTTTTTGAT	TTTGACTGCC	TCATCGTTTT
45	551	TGCCAAGCGG	GTTTTATACC	GACATCCTCA	ACCAACCCAA	TACTTATTAC
	601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	GCAGGTTGCG	TGCTGGCGGT
	651	TTACGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAAT	GGAAACCGGC
	701	AGTTGCTTTC	ATCACTCTGC	TTGCGGCGAT	TGCTTGCTGC	CCTGTTTCGT
	751	ATTGACAAAC	ACAATCCGTT	TATCCCGGGA	ATGACCTGCG	TCCTTCCCTG
50	801	CCTGCTGACG	GCACTGCTTA	TCCGGAGTAT	GCAATACGGG	ACACTTCCGA
	851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTGCGCAA	AATCTCTTAT
	901	TCCCTATACC	TGTACCATTT	GATTTTATTT	GCTTTCGCCC	ATTACATTAC
	951	AGGCGACAAA	CAGCTCGGAC	TGCCTGCCGT	ATCGGCGGTT	GCCGCGTTGA
	1001	CGGCCGGATT	TTCCCTGTTG	AGTTATTTAT	TGATTGAACA	GCCGCTTAGA
55	1051	AAACGGAAGA	TGACCTTCAA	AAAGGCATTT	TTCTGCCTCT	ATCTCGCCCC
	1101	GTCCCTGATA	CTTGTCGGTT	ACAACCTGTA	CGCAAGGGGG	ATATTGAAAC
	1151	AGGAACACCT	CCGCCCGTTG	CCCGGCGCGC	CCCTTGCTGC	GGAAATCAT

-454-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

1201 TTTCCGGAAC CCGTCCTGAC CCTCGGCGAC TCGCACGCCG GACACCTGAG  
1251 GGGGTTTCTG GATTATGTCG GCAGCCGGGA AGGGTGGAAG GCCAAATCC  
1301 TGTCCTCGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC  
1351 AACCCGTTAT GTCGAAAATA CCGGGATGAA GTTGAAAAAG CCGAAGCCGT  
1401 TTTTCATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCTGTGCCGA  
1451 GATTGAAGC GCAATCCTTC CTAATACCGG GGTTCACAGC CCGATTGAG  
1501 GAAACCGTCA AAAGGATAGC CGCCGTCAAA CCCGTCTATG TTTTTCGAAA  
1551 CAACACATCA ATCAGCCGTT CGCCCTGAG GGAGGAAAAA TTGAAAAGAT  
1601 TTGCCGCAAA CCAATATCTC CGCCCATTC AGGCTATGGG CGACATCGGC  
1651 AAGAGCAATC AGGCGGTCTT TGATTGATT AAAGATATTC CCAATGTGCA  
1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG  
1751 GCCGTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT  
1801 TATATGGGGG GGAATATCCA CAAACACGAA CGCCTGCTTA AATCTTCCCA  
1851 CGGCGGCCGA TTGCAGTAG

15 This corresponds to the amino acid sequence <SEQ ID 830; ORF128-1>:

1 MQAVRYRPEI DGLRAVAVL VMIFHLNNRW LPGGFLGVDI FFVISGFLIT  
51 GIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN  
101 QMRKTVELSA VFSLNIYLG FQQGYFDLSAD ENPVLHIWSL AVEEQYLLY  
151 PLLLIFFCKK TKSLRVLNRI SIILFLILTA SSFLPSCFYT DILNQPNYY  
201 LSTLRFPELL AGSLLAVYGO TQNGRRQTAN GKRQLLSSLC FGALLACLFV  
251 IDKHNPFIPG MTLPLCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY  
301 SLVLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSL SYYLIEQPLR  
351 KRKMTFKKAF FCLYLAPSLI LVGYNLYARG ILKQEHRLPL PGAPLAAENH  
401 FPETVLTGLD SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD  
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGCG PVPFPAQSF LIPGFARFR  
501 ETVKRIAIVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG  
551 KSNQAVFDLI KDIPNVHVDV AQKYLKNTV EIYGRYLYGD QDHLTYFGSY  
601 YMGREFHKHE RLLKSSHGGA LQ\*

Computer analysis of this amino acid sequence gave the following results:

30 Homology with hypothetical integral membrane protein HI0392 of *H.influenzae* (accession number U32723)

ORF128 and HI0392 show 52% aa identity in 180aa overlap:

35  
40  
45  
50  
55  
60

Orf128: 1 VSLASVIASQIFLYEDFNQMRKTVELSAVFSLNIYLG FQQGYFDLSADENPVLHIWSLAV 60  
++L S IAS IF+Y DEN++RKT+EL+ FLSN YLG QGYFDLSA+ENPVLHIWSLAV  
HI0392: 46 MALVSFIASAIIFYNDFNKLRTIELAIAFLSNFYLGITQGYFDLSANENPVLHIWSLAV 105  
Orf128: 61 EEQXXXXXXXXXIFCCKKTKSLRVLNRSIILFLILTASSFLPSGFYTDILNQPNYYLS 120  
E Q I KK + ++VL I++ILF IL A+SE+ + FY ++L+QPN YYLS  
HI0392: 106 EGQYYLIFPLILILAYKKFREVKVLFITLILFFILLATSFVSANFYKEVLHQPNIYYLS 165  
Orf128: 121 TLRFPELLAGSLLAVYGO TQNGRRQTANGKRQLLSSLCFGALLACLFVIDKHNPFIPGMT 180  
LRFPELL GSLLA+Y N + Q + +L+ L L +CLF+++ + FIPG+T  
HI0392: 166 NLRPELLVGSLLAIYHNLSN-KVQLSKQVNNILAILSTLLFSCFLMNNNIAFIPGIT 224

Homology with a predicted ORF from *N.meningitidis* (strain A)

45 ORF128 shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) from strain A of *N.meningitidis*:

50  
55  
60

orf128.pep  
orf128a  
orf128.pep  
orf128a  
orf128.pep

VSLASVIASQIFLYEDFNQMRKTVELSAVF  
ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVF  
LSNIYLG FQQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNRSI  
LSNIYLG FQQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNRSI  
ILFLILTASSFLPSGFYTDILNQPNYYLSTLRFPELLAGSLLAVYGO TQNGRRQTANGK

-455-

orf128a	ILFLILTATSEFLPSGFYTDILNQPNNTYYLSTLRFPELLAGSLLAVYGGTQNGRRQTANGK	180	190	200	210	220	230
5	orf128.pep	160	170	180	190	200	210
	orf128a	RQLSSSLCFGALLACLFVIDKHNPFI	PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI				
		240	250	260	270	280	290
10	orf128.pep	220	230	240			
	orf128a	VFVGKISYSLYLYHWIFIAFAPLIRGGKQLGLPA					
		300	310	320	330	340	350
15	orf128a	KMTFFKAFFCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTIGDSH	360	370	380	390	400
						410	

The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACGGATTGC	GGGCCGTCGC
51	CGTGCTATCC	GTCATGATTT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
101	GATTCCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC
151	GGCATCATT	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTTT	TCCGGGATTT
201	TTATACCCGC	AGGATTAAGC	GGATTTATCC	TGCTTTTATT	GCGGCCGTGT
251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
301	CAAATGCGGA	AAACCGTGGG	GCTTCTGCG	GTTTTCTTGT	CCAATATTTA
351	TCTGGGGTTT	CAGCAGGGGT	ATTCGATTT	GAGTGCCGAC	GAGAACCCCG
401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCCTGTAT
451	CCTCTTTTGC	TGATATTTTG	CTGCAAAAAA	ACAAAATCGC	TACGGGTGCT
501	GCGTAACATC	AGCATCATCC	TATTTCTGAT	TTTGACTGCC	ACATCGTTTT
551	TGCCAAGCGG	GTTTTATACC	GATATTCTCA	ACCAACCCAA	TACTTATTAC
601	CTTTGACAC	TGAGGTTTCC	CGAGCTGTTG	GCAGGTTCCG	TGCTGGCGGT
651	TTACGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAAT	GGAAAACGCG
701	AGTTGCTTTC	ATCACTCTGC	TTGCGCGCAT	TGCTTGCCGT	CCTGTTCTGT
751	ATTGACAAAC	ACAATCCGTT	TATCCCGGGA	ATGACCCTGC	TCCTTCCCTG
801	CCTGCTGACG	GCACTGCTTA	TCCGGAGTAT	GCAATACGGG	ACACTTCCGA
851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTCGGCAA	AATCTCTTAT
901	TCCCTATACC	TGTACCATTG	GATTTTATTT	GCTTTCGCCC	ATTACATTAC
951	AGGCGCAAAA	CAGCTCGGAC	TGCCTGCCGT	ATCGGCGGTT	GCCGCGTTGA
1001	CGGCCGGATT	TCCCTGTTG	AGTTATTATT	TGATTGAACA	GCCGCTTAGA
1051	AAACGGAAGA	TGACCTTCAA	AAAGGCATTT	TTCTGCCTCT	ATCTCGCCCC
1101	GTCCTTGATA	CTTGTCGGTT	ACAACCTGTA	CGCAAGGGGG	ATATTGAAAC
1151	PLTGAACACT	CCGCGCGTGG	CCCGCGCGCG	CCCTTGCTGC	GGAAAATCAT
1201	TTTCCGAAA	CCGTCCTGAC	CCTCGGCGAC	TCGCACGCCG	GACACCTGCG
1251	GGGGTTTCTG	GATTATGTCG	GCAGCCGGGA	AGGGTGGAAA	GCCAAAATCC
1301	TGTCCCTCGA	TCCGGAGTGT	TTGGTTTGGG	TAGATGAGAA	GCTGGCAGAC
1351	AACCCGTTAT	GTCGAAAATA	CCGGATGAA	GTTGAAAAAG	CCGAAGCCGT
1401	TTTCATTGCC	CAATTCTATG	ATTTGAGGAT	GGGCGGCCAG	CCCGTGCCGA
1451	GATTTGAAGC	GCAATCCTTC	CTAATACCCG	GGTTCCAGC	CCGATTCAGG
1501	GAAACCGTCA	AAAGGATAGC	CGCCGTCAA	CCCGTCTATG	TTTTTGCAAA
1551	CAACACATCA	ATCAGCCGTT	CGCCCTGAG	GGAGGAAAAA	TGAAAAAGAT
1601	TTGCCGCAAA	CCAATATCTC	CGCCCCATTC	AGGCTATGGG	CGACATCGGC
1651	AAGAGCAATC	AGGCGGTCTT	TGATTTGATT	AAAGATATTC	CCAATGTGCA
1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATATACG
1751	CGCGCTATCT	TTACGGCGAC	CAAGACCACC	TGACCTATTT	CGGTTCTTAT
1801	TATATGGGGC	GGGAATTTCA	CAAACACGAA	CGCCTGCTTA	AATCTTCTCG
1851	CGACGGCGCA	TTGCAGTAG			

This encodes a protein having amino acid sequence <SEQ ID 832>:

1	MQAVRYRPEI	DGLRAVAVLS	VMIFHLNNRW	LPGGFLGVDI	FFVISGFLIT
51	GIILSEIQNG	SFSFRDFYTR	RIKRIYPAFI	AAVSLASVIA	SQIFLYEDFN
101	QMRKTVELSA	VFLSNIYLG	QQGYFDLSAD	ENPVLHTWSL	AVEEQYLLY
151	PLLLIFCCKK	TKSLRVLRLNI	SIILFLILTA	TSFLEPSGFYT	DILNQPNNTYY
201	LSTLRFPELL	AGSLLAVYQG	TQNGRRQTAN	GKRQLSSSLC	FGALLACLFV
251	IDKHNPFI	PGMTLLLPCLLT	ALLIRSMQYG	TLPTIRLSAS	PIVFVGKISY
301	SLYLYHWIFI	FAFAHYITGDK	QLGLPAVSAV	AALTAGFSL	SYLLIEQPLR
351	KRMTEFKAF	FCLYLRASLI	LVGYNLYARG	ILKQEHRLPL	PGAPLAAENH
401	FPETVLTIGD	SHAGHLRGFL	DYVGSREGWK	AKILSLDSEC	LVWVDEKLAD
451	NPLCRKYRDE	VEKAEAVFIA	QFYDLRMGGQ	PVPRFEAQSF	LIPGFPARFR

501 ETVKRIA AVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG  
 551 KSNQAVFDLI KDIPNVHWVD AQKYL PKNTV EIYGRYLYGD QDHLTYFGSY  
 601 YMGREFHKHE RLLKSSRDGA LQ\*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5	orf128a.pep	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
	orf128-1	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
10	orf128a.pep	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128-1	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128a.pep	QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
15	orf128-1	QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
	orf128a.pep	TSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKROLLSSIC
	orf128-1	SSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKROLLSSIC
20	orf128a.pep	FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
	orf128-1	FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
25	orf128a.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLSYLLIEQPLRKRKMTFKKAF
	orf128-1	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLSYLLIEQPLRKRKMTFKKAF
30	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTGDSHAGHLRGFL
	orf128-1	FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTGDSHAGHLRGFL
	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
35	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128a.pep	PVPRFEAQSFILPGFPPARFRETVKRIA AVKPVYVFANNTSISRSPLREEKLRFAANQYL
	orf128-1	PVPRFEAQSFILPGFPPARFRETVKRIA AVKPVYVFANNTSISRSPLREEKLRFAANQYL
40	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYL PKNTVEIYGRYLYGDQDHLTYFGSY
45	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
	orf128-1	YMGREFHKHERLLKSSHGALQX

#### Homology with a predicted ORF from *N. gonorrhoeae*

50 ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N. gonorrhoeae*:

	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
55	orf128.pep	LSNIYLGFGQQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISI	90
	orf128ng	LSNIYLGFRGLGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCYKTKSLRVLNISI	172
60	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK	150
	orf128ng	ILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPPELLVGSLLAVYGQTQNGRRQTENGK	232
65	orf128.pep	RQLSSLLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	210
	orf128ng	RQLSLLCFGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPI	292

orf128.pep VFVGKISYSLYLHWIFIAFAPLIRGGKQLGLPA 244  
 |||||  
 orf128ng VFVGKISYSLYLHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLIEQPLRKR 352

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

1 ATGCAAGCTG TCCGATACAG GCCTGAAATT GACGGATTGC GGGCCGTCGC  
 51 CGTGCTATCC GTCATTATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG  
 101 GATTCCCTGGG GGTGGACATT TTCTTTGTCA TCTCGGGATT CCTCATTACC  
 151 AACATCATT TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATT  
 201 TTATACCCGC AGGATTAAGC GGATTATCC TGCTTTTATT GCGGCCGTGT  
 251 CCCTGGCTTC GGTGATTGCT TCTCAAATCT TCCTTTACGA AGATTTC AAC  
 301 CAAATGAGGA AAACCATAGA GCTTTCTACG GTTTTCTGT CCAATATTTA  
 351 TTTGGGGTTC CGATTGGGGT ATTTGCGATT GAGTGCCGAC GAGAACCCCG  
 401 TACTGCATAT CTGGTCTTTG GCGGTAGAGG AACAGTATTA CCTCCTGTAT  
 15 451 CCTCTTTTGC TGATATTCTG TTACAAAAA ACCAAATCAC TACGGGTGCT  
 501 GCGTAATATC AGCATCATCC TGTTCTGAT TTTGACCGCA TCATCGTTTT  
 551 TGCCGCGCGG GTTTTATACC GACATCCTCA ACCAACCCaa TACTATTAC  
 601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GTGGGTTCGC TGTGGCGGT  
 651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGAAAAT GGAAAACGGC  
 20 701 AGTTGCTTTC ATTACTCTGT TTCGGCGCat tgCTTGTCTG CCTGTTCCGT  
 751 ATCGACAAAC ACGATCCGTT TATCCCGGGA ATAACCCTGC TCCTTCCCTG  
 801 CTTGCTGACG GCGCTGCTTA TCCGGAGTAT GCAATACGGG ACACTTCCGA  
 851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT  
 901 TCCCTATACC TGTACCATTG GATTTTTATT GCCTTCGCCC ATTACATTAC  
 25 951 AGGCGACAAA CAGCTCGGAC TGCTGCGGT ATCGGCGGT GCCGCGTTGA  
 1001 CGGCCGATT TTCCCTGTTG AGCTATTATT TGATTGAACA GCCGCTTAGA  
 1051 AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTTT ATCTCGCCCC  
 1101 GTCCCTGATG CTGTGTCGGT ACAACCTGTA TTCAAGAGG ATATTGAAAC  
 1151 AGGAACACCT CCGCCCGCTG CCCGGCACGC CCGTTGCTGC GGAAAATAAT  
 30 1201 TTTCCGAAA CCGTCTTGAC CCTCGGCGAC TCGCACGCCG GACACCTGCG  
 1251 GGGGTTTCTG GATTATGTCG GCGGCAGGGA AGGTGGAAG GCTAAAATCC  
 1301 TGTCCCTCGA TTCGGAGTGT TTGGTTGGG TGGATGAGAA GCTGGCAGAC  
 1351 AATCCGTTGT GCCGAAAATA CCGGATGAA GTTGAAAAAG CCGAAGCTGT  
 1401 TTTGCTATGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCGTGCCCGA  
 35 1451 GATTTGAAGC GCAATCCTTC CTGATACCCG GGTTCAAAGC CCGATTGAGG  
 1501 GAAACCGTCA AGAGGATAGC CGCCGTCAAA CCTGTATATG TTTTGTGAAA  
 1551 CAATACATCA ATCAGCCGTT CTCCCTTGAG GGAGGAAAAA TTGAAAAGAT  
 1601 TTTGCTATAA CCAATACCTC CGGCCTATTC GGGCTATGGG GCACATCGGC  
 40 1651 AAGAGCAATC AGGCGGTCTT TGATTGGTT AAAGATATTC CCAATGTGCA  
 1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATACACG  
 1751 GACGCTATCT TTACGGCGAC CAAGACCACC TGACCTATT CGGTTCCTAT  
 1801 TATATGGGGC GGGAATTTCA CAAACACGAA CGCCTGCTCA AGCATTCGCC  
 1851 AGGCGGCGCA TTGCAGTAG

This encodes a protein having amino acid sequence <SEQ ID 834>:

45 1 MQAVRYRPEI DGLRAVAVLS VILFHLNLRW LPGGFLGVDI FFVISGFLIT  
 51 NIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN  
 101 QMRKTIELST VFLSNIYLG FRLGYFDLSAD ENPVLHIWSL AVEEQYLLY  
 151 PLLLIFYCYK TKSLRVLRLNI SIILFLILTA SSFLPAGFYT DILNQPNITY  
 201 LSTLRFPPELL VGSLLAVYGQ TQNGRRQ TEN GKRQLLSLLC FGALLVCLFV  
 50 251 IDKHDPPFIPG ITLLPLCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY  
 301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSL SYLIEQPLR  
 351 KRKMTFKKAF FCLYLAPSLM LVGYNLYSRG ILKQEHRLRPL PGTFVAEENN  
 401 FEETVLTLD SHAGHLRGFL DYVGGREGWK AKILSLDSEC LVWVDEKLD  
 451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFKARFR  
 55 501 ETVKRIAAVK PVYVFANNTS ISRSPLREEK LKRFAINQYL RPIRAMGDTG  
 551 KSNQAVFDLV KDIPNVHVD AQKYLPKNTV EIHGRYLYGD QDHLTYFGSY  
 601 YMGREFHKHE RLLKHSRGA LQ\*

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

60 orf128-1.pep MQAVRYRPEIDGLRAVAVLSVMIFHLNLRWLPGGFLGVDIFFVISGFLITGIILSEIQNG  
 |||||  
 orf128ng MQAVRYRPEIDGLRAVAVLSVIFHLNLRWLPGGFLGVDIFFVISGFLITNIILSEIQNG  
 orf128-1.pep SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG  
 |||||  
 65 orf128ng SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVFLSNIYLG

-458-

```

5  orf128-1.pep  QQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLNRNISIILFLILTA
   orf128ng     RLGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCYKKTSLRVLNRNISIILFLILTA

10 orf128-1.pep  SSFLPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLLSLLC
   orf128ng     SSFLPAGFYTDILNQPNNTYYLSTLRFPPELLVGSLLAVYGQTQNGRRQTENGKRQLLSLLC

15 orf128-1.pep  FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASP V FVGKISY
   orf128ng     FGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASP V FVGKISY

20 orf128-1.pep  SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
   orf128ng     SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF

25 orf128-1.pep  FCLYLAPSLILVGNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLDSSHAGHLRGFL
   orf128ng     FCLYLAPSLMLVGNLYSRGILKQEHRLPLPGTPVAAENHFPETVLTGLDSSHAGHLRGFL

30 orf128-1.pep  DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
   orf128ng     DYVGGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ

35 orf128-1.pep  PVPREFEASFLIPGFPAFRET VKRIA AVKPVYVFANNTSISRSPLREEKLRFAANQYL
   orf128ng     PVPREFEASFLIPGFKA FRET VKRIA AVKPVYVFANNTSISRSPLREEKLRFAINQYL

40 orf128-1.pep  RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYL PKNTVEIYGRYLYGQDHLTYFGSY
   orf128ng     RPIRAMGDIGKSNQAVFDLVKIDIPNVHWVDAQKYL PKNTVEIHGRYLYGQDHLTYFGSY

45 orf128-1.pep  YMGREFHKHERLLKSSHGGALQX
   orf128ng     YMGREFHKHERLLKHSRGGALQX
                        610      620

```

In addition, ORF218ng shows homology to a hypothetical *H.influenzae* protein:

```

40 sp|P43993|Y392_HAEIN  HYPOTHETICAL PROTEIN HI0392 >gi|1074385|pir||B64007
   hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
   >gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus
   influenzae] Length = 245
   Score = 239 bits (604), Expect = 3e-62
   Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)

45 Query: 38  VDIFFVISGFLITNIILSEIQNGSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXXXXXFLYE 97
   +DIFFVISGFLIT II++EIQ SFS + FYTRRIKRIYP F+Y
   Sbjct: 1  MDIFFVISGFLITGIIITEIQNSFSLKQFYTRRIKRIYPAFITVMALVSFIASAIIFIYN 60

50 Query: 98  DFNQMRKTIELSTVFLSNIYLGFR LGYFDLSADENPVLHIWSLAVEEQXXXXXXXXXIFC 157
   DFN++RKTIEL+ FLSN YLG GYFDLSA+ENPVLHIWSLAVE Q I
   Sbjct: 61  DFNKLRKTIELAIAFLSNFYLG LTQGYFDLSANENPVLHIWSLAVEGQYYLIFPLILILA 120

55 Query: 158 YKKTSLRVLNRNISIILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPPELLVGSLLAV 217
   YKK + ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+
   Sbjct: 121 YKKFREV KVLFIITLILFFILLATS FVSANFYKEVLHQPNIIYLSNLRFPELLVGSLLAI 180

60 Query: 218 YGQTQNGRRQTENGKRQLLSLLCFGALLVCLFVIDKHDPFIPGIT 262
   Y N + Q +L++L L CLF+++ + FIPGIT
   Sbjct: 181 YHNLSN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

```

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 99**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51  GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATTG GGTCTGTTGT
5  101  TGGCGTTGGC GCGCCTGATT CACTTGAAAA AAGCCGGTGC GCCGATGCGC
151  GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201  CCGGGGTACG CCGCTGTTTG TGCAGATTGT GATTGGGCGG TATGTGTGGT
251  TTCCGTTTTT CGTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

10  1  ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
51  VLAWALRKVS LLYVTLFRGT PLFVQIVIWA YVWFPPFV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTTC GTTTGGACAT TATTACGAA TACCGCTGGA TGTTTCTTTA
51  CCGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT
15  101  CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTGGAAAAAA
151  GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AAGTTTCGCT
201  GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTT CAGATTGTGA
251  TTTGGGCGTA TGTGTGGTTT CCGTTTTCG TCCATCCTTC AGACGGCATT
301  TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
20  351  GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401  AGATTTTCCG CCGCGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
451  GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
501  GCCGCAGGCA TTGCGCCGCA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
551  CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGGAGTTG
25  601  GCGTATGTTC AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
651  GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701  GGATATTCTT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

1  MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
30  51  AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVIWAYVWF PFFVHPSDGI
101  LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151  ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL
201  AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

Computer analysis of this amino acid sequence gave the following results:

### 35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.meningitidis*:

```

10  20  30  40  50
orfl29.pep  IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
40  orfl29a  MDFFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
10  20  30  40  50  60
60  70  80
orfl29.pep  ALRKVSLLYVTLFRGTPLEFVQIVIWAYVWFPPFV
45  orfl29a  ALRKVSLLYVTLFRGTPLEFVQIVIWAYVWFPPFVHPSDGILVSGEAAIALRRYGPLIAG
70  80  90  100  110  120
50  orfl29a  SLALIANSQAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPLAS
130  140  150  160  170  180

```

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

```

1  ATGGATTTTC GTTTGGACAT TATTACGAA TACCGCTGGA TGTTTCTTTA
51  CCGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCGACG GCGGGCGGTT

```

```

5      101  CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAGAAA
      151  GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT
      201  GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
      251  TTTGGGCGTA TGTGTGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
      301  TTGGTTAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
      351  GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
      401  AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
      451  GCGCGTCTCT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
      501  GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
10     551  CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GGCGGAGTTG
      601  GCGTATGTTC AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
      651  GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
      701  GGATATTCTT GCGTTTGGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 840>:

```

15      1  MDFRFDIIE YRWMFLY GAL TTLGLTVVAT AGGSVLG LLL ALARLIHLEK
      51  AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVIWAYVWF PFFVHPSDGI
      101  LVSGEAAIAL RRGYGPLIAG SLALIANS GA YICEIFRAGI QSIDKGQMEA
      151  ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
      201  AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

20 ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

```

      orf129a.pep  MDFRFDIIEYRWMFLYGALTTLGLTVVATAGGSVLG LLLALARLIHLEKAGAPMRVLAW
      orf129-1     MDFRFDIIEYRWMFLYGALTTLGLTVVATAGGSVLG LLLALARLIHLEKAGAPMRVLAW

25     orf129a.pep  ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFVHPSDGI LVSGEAAIALRRGYGPLIAG
      orf129-1     ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFVHPSDGI LVSGEAAIALRRGYGPLIAG

30     orf129a.pep  SLALIANS GAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPPLAS
      orf129-1     SLALIANS GAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPPLAS

      orf129a.pep  EFITLLKDSS LLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
35     orf129-1     EFITLLKDSS LLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE

      orf129a.pep  KRYNPQHRX
      orf129-1     KRYNPQHRX
40

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from *N.gonorrhoeae*:

```

45     orf129.pep      IIYEYRWMFLYGALTTLGLTVVAXAGGSVLG LLLALARLIHLEKAGAPMRVLAW      54
      orf129ng        MDFFRFDIIEYRWMFLYGALTTLGLTVVATAGGSVLG LLLALARLIHLEKAGAPMRVLAW      60

      orf129.pep      ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFVFFV      88
      orf129ng        ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFVFFVILHTAFLGNAMRQSRVPDKGRWIAG      120

```

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

```

55      1  MDFRFDIIE YRWMFLY GAL TTLGLTVVAT AGGSVLG LLL ALARLIHLEK
      51  AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVIWAYVWF PFFVILHTAF
      101  LGNAMRQSRV VPDKGRWIAG SLELNCQPRG RKTRGEFPPG ESNLGTEPRN
      151  PLSMGQRREF GCENWYPPQN FIKK*

```

Further work revealed the following gonococcal sequence <SEQ ID 843>:

```

1  ATGGATTTTc gTTTTGACAT TATTTAcgaA TACCGCTGGA TGTTTCTTTA

```



-461-

```

5      51  CGGCGCACTG  Acgaccttgg  ggctgacggt  cgtggcgacg  gCGGGCGGTT
      101  CGGtattggG  TCTGTTGTTG  GCGTTGGCGC  GCCTGATTCA  CTGGAAGAAA
      151  GCCGGTGCGC  CGATGCGCGT  GCTGGCGTGG  GCGTTGCGTA  AGGTTTCGCT
      201  GCTGTACGTT  ACCCTGTTCC  GGGGTACGCC  GCTGTTTGTG  CAGATTGTGA
      251  TTTGGGCGTA  TGTGTGGTTT  CCGTTTTTCG  TCCATCCTTC  AGACGGCATT
      301  TTGGTCAGCG  GCGAGGCGGC  AATCGCGCTG  CGTCGCGGAT  ACGGGCCGCT
      351  GATTGCCGGT  TCTTTGGCAC  TGATCGCCAA  CTCGGGGGCG  TATATCTGTG
      401  AGATTTTCCG  CGCGGGCATC  CAGTCTATAG  ACAAAGGACA  GATGGAGGCG
      451  GCGTGTTCCT  TGGGACTGAC  CTATCCGCAG  GCGATGCGCT  ATGTGATTCT
      501  GCCGCAGGCA  TTGCGCCGTA  TGCTGCCGCC  TTTGGCGAGC  GAGTTCATCA
      551  CGCTCTTGAA  AGACAGCTCG  CTGCTGTCGG  TCATTGCTGT  GCGGGAGTTG
      601  GCGTATGTTT  AGAATACGAT  TACGGGCCGG  TATTCGGTTT  ATGAAGAACC
      651  GCTTTACACC  GCCGCCCTGA  TTTATCTGTT  GATGACGACT  TTCTAGGCTT
      701  GGATATTCTT  GCGTTTGGA  AAACGTTACA  ATCCGCAACA  CCGCTGA

```

15 This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

```

      1  MDFRFDIIYE  YRWMFLYGAL  TTLGLTVVAT  AGGSVLGLLL  ALARLIHLEK
      51  AGAPMRVLAW  ALRKVSLLYV  TLFRTPLFV  QIVIWAYVWF  PFFVHPSDGI
      101  LVSGEAAIAL  RRGYGPLIAG  SLALIANSQA  YICEIFRAGI  QSIDKGQMEA
      151  ARSLGLTYPO  AMRYVILPQA  LRRMLPLAS  EFITLLKDSS  LLSVIAVAEL
      201  AYVQNTITGR  YSVYEEPLYT  VALIYLLMTT  FLGWIFLRLE  KRYNPQHR*

```

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

```

      orf129-1.pep  MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      orf129ng-1    MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      25
      orf129-1.pep  ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFFFVHPSDGIILVSGEAAIALRRGYGPLIAG
      orf129ng-1    ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFFFFVHPSDGIILVSGEAAIALRRGYGPLIAG
      30
      orf129-1.pep  SLALIANSQAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLAS
      orf129ng-1    SLALIANSQAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLAS
      35
      orf129-1.pep  EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
      orf129ng-1    EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
      40
      orf129-1.pep  KRYNPQHRX
      orf129ng-1    KRYNPQHRX

```

In addition, ORF129ng-1 is homologous to an ABC transporter from *A.fulgidus*:

```

2650409(AE001090) glutamine ABC transporter, permease protein (glnP)
[Archaeoglobus fulgidus]Length = 224
Score = 132 bits (329), Expect = 2e-30
45  Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)

Query: 65  VSLLYVTLFRGTPLFVQIVIWAYVWFFFFVHPSDGIILVSGEAAIALRRGYGPLIAGSLAL 124
      +S  YV + RGTPV VQI+I      +F  P+ GI +  E  A      G +AL
Sbjct: 58  ISTAYVEVIRGTPLLVQILI-----VYFGLPAIGINLQPEPA-----GIIAL 99

Query: 125  IANSQAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLASEFIT 184
      SGAYI EI RAGI+SI  QMEAA SLG+TY  QAMRYVI  PQA R +LP L +EFI
Sbjct: 100  SICSGAYIAEIVRAGIESIPIGQMEAAARSLGMYLQAMRYVIFPQAFRNILPALGNEFIA 159

55  Query: 185  LLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAAALIYLLMTTFLGWIFLRLEKR 242
      LLKDSSLLSVI++ EL  V  I      P  AL YL+MT  L  +  +K+
Sbjct: 160  LLKDSSLLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLMMTIPLSRLVAYSQKK 217

```

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful  
 60 antigens for vaccines or diagnostics, or for raising antibodies.

**Example 100**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1   ..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA
5  51  TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT
    101  GGCTGCCCGC GCAAACCGCC GGTTTACC GGTTCGCCGT CGGCTTCATC
    151  CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACCTCT TACGTAAACA
    201  cTACGTCCGC ACTTATTACy TGCTCCAACCT CTTTGCCGCC GCAGgcTAgT
    251  TTGTGGACAG GCGCGGCGwA ATTACAAAC CTGCCCGCyT CCGCGCCCTT
    301  GCACCTGATT ACCCTCGCG GCATGATGGG CGGCGTGATG ATGGTGTGGc
10  351  TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAACTCGA CTACCCCAAA
    401  CTCTGCCGCA TTGCCGTCCT CATCCTTTTC GCCGCCGCCG TCTCGCGCGC
    451  TTTCTTGrTG AACGTGAACC CGrTATTTT CATTACCGTT CCTGCGATTCT
    501  TGACCGCCGC CGTATTCGTA CTGTATCTT TCrCGTTTAT ACCGATATTT
    551  CGGGCGAATG CGTTTACAGA CGATCCGGAr TAr

```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1   ..LKECRLKDPV FIPNIVYKNI AITFLLLHAA AELWLPAQTA GFTALAVGFI
51  LLAKLRELHH HELLRKHYVR TYLLQLFAA AGSLWTGAAX LQNLPAAPL
101 HLITLGGMMG GVMVWLTA LWSHGFTKLD YPKLCRIAVP ILFAAAVSRA
151 FLXNVNEXFF ITVPAILTAA VFVLYLFXFI PIFRANAFTD DPE*

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1   ATGCGGCCGT TTTTCGTCGG CGCGGCGGTG CTGCCCAC TCGGTGCGCT
51  GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
101 TGGAACTTAT GCTGCCGGCG GCATACGGCG GTTTTTTGAC TGCGGCTTTG
25  151  TTGGACTGGA CGGTTTTTTC GGGTAACCTG AAACCTGTCG CGACTTTGAT
    201  GGCGGCATTA TTGCTCGCCG CATCCGCTAT ACTGCCCTTT TCGCCGCAAA
    251  CTGCCTCGTT TTTTCGTCGCC GCCTATTGGC TGGTGTGCT GCTGTTCTGC
    301  GCCCGGCTGA TTTGGCTAGA CCGAAACACC GACAACTTCG CCCTGCTAAT
    351  GTTACTTGCC GCGTTCACCTG TTTTTCAGAC GGCATATGCC GTCAGCGGCG
    401  ATTTGAACCT GTTGCGCGCG CAAAGTCATC TAAATATGGC GGCGGTGATG
30  451  TTCGTATCCG TCGCGCTCAG TATTCTTTTG GGCGCGGAAG CCCTGAAAGA
    501  ATGCCGCTCG AAAGACCTG TTTTATTCC AAATATCGTT TATAAAAAACA
    551  TCGCCATTAC TTTCTGCTC TTGCACGCCG CCGCCGAAC TTTGGCTGCC
    601  GCGCAAACCG CCGTTTTTAC CGCGCTCGCC GTCGGCTTCA TCCTGCTCGC
    651  CAAGCTGCGT GAGCTTCACC ATCACGAACT CTTACGTAAA CACTACGTCC
35  701  GCACTTATTA CCTGCTCCAA CTCTTGCCG CCGCAGGCTA TTTGTGGACA
    751  GGCGCGGCGA AATTACAAAA CCTGCCCGCC TCCGCGCCCC TGCACCTGAT
    801  TACCCTCGGC GGCATGATGG GCGGCGTGAT GATGGTGTGG CTGACCGCCG
    851  GACTGTGGCA CAGCGGCTTT ACCAACTCG ACTACCCCAA ACTCTGCCCG
    901  ATTGCCGTCC CCATCCTTTT CGCGCGCGCC GTCTCGCGCG CTTTCTTGAT
40  951  GAACGTGAAC CCGATATTTT TCATTACCGT TCCTGCGATT CTGACCGCCG
    1001 CCGTATTCGT ACTGTATCTT TTCACGTTA TACCGATATT TCGGGCGAAT
    1051 GCGTTTACAG ACGATCCGGA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1   MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
45  51  LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFVVA AYWLVLFFF
    101  ARLIWLDRNT DNFAILLMLLA AFTVFOTAYA VSGDLNLLRA QVHLNMAAVM
    151  FVSVRVSILL GAEALKECRL KDPVFIPNIV YKNIAITFL LHAALWLP
    201  AQTAGFTALA VGFILLAKLR ELHHHELLRK HYVRTYLLQ LFAGAGYLWT
    251  GAAKLQNLPA SAPHLITLG GMMGVMMVW LTAGLWHS GF TKLDYFKLCR
50  301  IAVPILFAAA VSRFLMNVN PIFFITVPAI LTAAVFVLYL FTFIPIFRAN
    351  AFTDDPE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N.*

55 *meningitidis*:

25	1	ATGCGGCGGT	TTTTCGTCGG	CGCGGCGGTG	CTTGCCATAC	TCGGTGCCT
	51	GGTGTTTTTC	ATCAACCCCG	GTGCCATCGT	CCTGCACCGC	CAAATTTTCT
	101	TGGAACCTTAT	GCTGCCGGCG	GCATACGGCG	GTTTTTTTGAC	TGCGGCTTTG
	151	TTGGACTGGA	CGGGTTTTTC	GGGTAACCTG	AAACCTGTCG	CGACTTTGAT
	201	GGCGGCATTA	TTGCTCGCCG	CATCCGCTAT	CACCTGCTTT	TCGCCGCAAA
30	251	CTGCCTCGTT	TTTCGTGCGC	GCCTATTGGC	TGGTGTTGCT	GCTGTTCTGC
	301	GCCCCGCTGA	TTTGGCTAGA	CCGAAACACC	GACAACTTGC	CCCTGCTAAT
	351	GTTACTTGCC	GCGTTCACTG	TTTTTCAGAC	GGCATATGCC	GTGAGCGGCG
	401	ATTTGAACCT	GTTGCGCGCG	CAAGTGCATC	TAAATATGGC	GGCGGTGATG
	451	TTCGTATCCG	TGCGCGTCAG	TATTCTTTTG	GGCGCGGAAG	CCCTGAAAGA
35	501	ATGCCGTCTG	AAAGACCCAG	TATTCATCCC	CAATGTCGTC	TATAAAAAACA
	551	TCGCCATTAC	CTTCTGTCTC	CTGCACGCCG	CGCGCGCACT	TTGGCTGCCT
	601	GCGCAAAACCG	CCGGTTTTAC	CTCGCTCGCC	CTCGGCTTTA	TCCTGCTTGC
	651	CAAGCTGCGT	GAGCTTCACC	ATCACGAACT	CCTGCGCAAA	CACTACGTCC
	701	GCACTTATTA	CCTGCTCCAA	CTCTTTGCGG	CCGAGGCTA	TTTGTGGACA
40	751	GGCGCGGCGA	AATTACAAAA	CCTGCCCGCC	TCCGCGCCCC	TGCACCTGAT
	801	TACCTTCGGT	GGCATGATGG	GCAGCTGTAG	GATGGTGTGG	CTGACTGCCG
	851	GACTGTGGCA	CAGCGGCTTT	ACCAAGCTCG	ACTACCCGAA	ACTCTGCCGC
	901	ATCGCCGTCC	CCATCCTNTT	CGCCGCCGCC	GTTTCGCGCG	CTGTTTTAAT
	951	GAACGTAAAC	CCGATATTCT	TCATCACCGT	CCCCGCAATT	CTGACGCGCG
45	1001	CCGTGTTTCGT	GCTTTACCTG	CTGACATTCT	TACCGATCTT	TCGGGCGAAC
	1051	GCGTTTACAG	ACGATCCGGA	ATAA		

50

1	MRPFFVGA	AV	LAILGALV	FF	INPGAIVL	LHR	QIFLELM	LPLA	AYGGFL	TAAL
51	LDWTGFS	GNL	KPVATLM	AAL	LLAASAIL	PF	SPQTASF	FFVA	AYWLVL	LLFC
101	ARLVLD	RNT	DNFALL	MLLA	AFTVFQ	TAYA	VSGDNL	LLLR	QVHLNM	AAVM
151	FVSVRV	SILL	GAEALKE	CR	L	KDPVFI	PNVV	YKNIAT	FTFL	LHAAAE
201	AQTAGF	TSLA	VGFI	LAKL	R	ELHHHE	LLRK	HYVRTY	YLLQ	LFAAAG
251	GAAKLQ	NLPA	SAPLHL	ITLG	GMMGSV	MMVW	LTAGLW	HSGF	TKLDY	PKLCR
301	IAPVIL	FAAA	VSRV	LMNVN	PIFFIT	VPAI	LTAAEV	FLYL	LTFVPI	FRAN
351	AFTD	DP	E	*						

55

```

60      orf130a.pep      MRPFFVGAAVLAILGALVEFFINPGAIVLHRQIFLELMPLAAYGGFLTAALLDWTGFSGNL
      |
      |
      |
      orf130-1         MRPFFVGAAVLAILGALVEFFINPGAIVLHRQIFLELMPLAAYGGFLTAALLDWTGFSGNL
      |
      |
      |
      orf130a.pep      KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVLLLFCARLIWLDNRNTDNFALLMLLA
      |
      |
      |
      orf130-1         KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVLLLFCARLIWLDNRNTDNFALLMLLA
      |
      |
      |
65      orf130a.pep      AFTVFQTAAYAVSGDLNLLRAOVHNLMAAVMFVSVRVSIILGAEALKECRLKDPVFIPNVV

```

```

      orf130-1      |||||
      orf130a.pep  YKNIAITFLLHAAELWLPAGTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLQ
5      orf130-1      YKNIAITFLLHAAELWLPAGTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQ
      orf130a.pep  LFAAAGYLWTGA AKLQNL PASAPLHLITLGGMMGSMVMVWLTAGLWHS GFTKLDY PKLCR
10     orf130-1      LFAAAGYLWTGA AKLQNL PASAPLHLITLGGMMGSMVMVWLTAGLWHS GFTKLDY PKLCR
      orf130a.pep  IAVPILFAAAVSRVLMNVNPIFFITVPAILTA AAVFVLYLLTFVPIFRANAFTDDPE
15     orf130-1      IAVPILFAAAVSRVLMNVNPIFFITVPAILTA AAVFVLYLLTFVPIFRANAFTDDPE

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

```

20     orf130.pep      LKECRLKDPVFIPNIVYKNIAITFLLHAA 30
      orf130ng      LNLRAQVHLNMAAVMFVSVRVSVLLGTETLKECRLKDPVFIPNVIYKNIAIT-LLHAA 201
      orf130.pep      AELWLPAGTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX 90
25     orf130ng      AELWLPAGTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGA AK 261
      orf130.pep      LQNL PASAPLHLITLGGMMGSMVMVWLTAGLWHS GFTKLDY PKLCRIAVPILFAAAVSR 150
30     orf130ng      LQNL PASAPLHLITLGGMTGGVMVWLTAGLWHS GFTKLDY PKLCRIAVSILFASAVSR 321
      orf130.pep      FLXNVNPFITVPAILTA AAVFVLYLFXFIPIFRANAFTDDPE 193
      orf130ng      VLMNVNPIFFITVPEILTA AAVFVLYLLTFVPIFRANAFTDDPE 364

```

35 An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino acid sequence <SEQ ID 852>:

```

1  MNKFFTHPMR PFFVGA AVLA ILGALVFFHQ PRRYHEAPPN FLGTYAAGCI
51 RRFDDYRFVG PDGFFRQ PET CRYFDGGVVA CCGCFIAVET ATCRIFRRRL
101 LAGVA AVLRL ADLARRQHRT LRSVDVTA AF TVFQTAYAVS GDLNLLRAQV
40 151 HLNMAAVMFV SVRVSVLLGT ETLKECRLKD PVFIPNVIYK NIAITLLLHA
201 AAELWLPAGT AGFTALAVGF ILLAKLRELH HHELLRKHYV RTYLLQLFA
251 AAGYLWTGAA KLQNL PASAP LHLITLGGMT GGVMVWLT A GLWHS GFTKL
301 DYPKLCRIAV SILFASAVSR AVL MNVNPIF FITVPEILTA AVFMYLLTF
351 VPIFRANAFT DDPE*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

```

45 1  ATGCGCCCGT TTTTCGTCGG TGCGGCAGTA CTTGCCATAC TCGGTGCGTT
      51  GGTGTTTTTT ATCAACCCCG GCGCTATCAT CCTGCACCGC CAAATTTTCT
101 101 TGGAACTTAT GCTGCCGGCT GCATACGGCG GTTTTTTGAC TACCGCTTGT
      151  TTGGACCGGA CGGGTTTTTC AGGCAACCTG AAACCTGCCG CTACTTTGAT
50 201  GCGCGTGTTG TTGCTTGTTG CGGCTGTTTT ATTGCCGTTT TTACCGCAAC
      251  TTGCCGCATT TTTTCGTCGCC GCCTATTGGC TGGTGTGCT GCTGTTCTGC
301 301 GCCTGGCTGA TTTGGCTCGA CCGCAACACC GACAACCTCG CTCTGTTGAT
      351  GTTACTTGCC GCATTTACCG TTTTTCAGAC GGCCTATGCC GTCAGCGGCG
401 401 ATTTGAACTT ACTGCGCGCG CAAGTGCATT TGAATATGCC GGCGGTCATG
      451  TTCGTATCCG TCCGCGTCAG CGTCCTTTTG GGCACGGAAA CCCTGAAAGA
55 501  ATGCCGCTCG AAAGACCCCG TATTCATCCC CAACGTTATC TATAAAAACA
      551  TCGCCATCAC CCTGCTGCTG CACGCCGCGC CCGAACTTTG GCTGCCGCGC
601 601 CAAACCGCGG GTTTTACTGC GCTTGCCGTC GGCTTCATCC TGCTCGCCAA
      651  GCTGCGCGAA CTGCACCATC ACGAACTCTT ACGCAAACAC TACGTCCGCA
701 701 CTTATTACCT GCTCCAGCTC TTTGCCGCGC CAGGTATCT GTGGACAGGC
      751  GCGGCGAAAC TGCAAAACCT GCCCGCCTCC GCGCCCTGC ACCTGATTAC
801 801 CCTCGCGCGC ATGACGGGTG GCGTGATGAT GGTGTGGCTG ACTGCCGGAC
851 851 TGTGGCACAG CGGCTTTACC AAACCTCGACT ACCCGAACT CTGCCGCATC

```

-465-

```

901  GCCGTCTCCA TCCTTTTCGC CTCCGCCGTT TCGCGCGCTG TTTAATGAA
951  CGTGAATCCG ATATTCTTCA TCACCGTTCC CGAGATTCTG ACCGCCGCCG
1001 TGTTCATGCT TTACCTGCTG ACGTTCGTAC CGATTTTTCG AGCGAACGCG
1051 TTTACAGACG ATCCGGAATA A

```

5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

```

1  MRPFFVGA AV LAILGALVFF INPGAII LHR QIFLELMLPA AYGGFLTTAL
51  LDRTGFSGNL KPAATLMAVL LLVAAVLLPF LPQLAAFFVA AYWLVL LLLFC
101 AWLIWLD RNT DNFALLMLLA AFTVFQ TAYA VSGDLNLLRA QVHLNMAAVM
151 FVSVRVSVLL GTETLKECRL KDPVFIPNVI YKNIAITLLL HAAAE LWLPA
10  201 QTAGFTALAV GFILLAKLRE LHHHELLRKH YVRTYYLLQL FAAAGYLWTG
251 AAKLQNL PAS APLHLITLGG MTGGVMMVWL TAGLWHS GFT KLDY PKLCRI
301 AVSILFASAV SRAVLMNVNP IFFITVPEIL TAAVFMLYLL TFVPIFRANA
351 FTDDPE*

```

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

```

15  orf130-1.pep  MRPFFVGA AVLAILGALVFFINPGAIVLHRQIFLELMLPAAYGGFLTAALLDWTGFSGNL
    orf130ng-1   MRPFFVGA AVLAILGALVFFINPGAII LHRQIFLELMLPAAYGGFLTALLDRTGFSGNL

20  orf130-1.pep  KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LLLFCARLIWLD RNTDNFALLMLLA
    orf130ng-1   KPAATLMAV LLLVAAVLLPFLPQLAAFFVAAYWLVL LLLFCAWLIWLD RNTDNFALLMLLA

25  orf130-1.pep  AFTVFQ TAYAVSGDLNLLRAQVHLNMAAVMFVSVRSILLGAEALKECRLKDPVFIPNIV
    orf130ng-1   AFTVFQ TAYAVSGDLNLLRAQVHLNMAAVMFVSVRSVLLGTETLKECRLKDPVFIPNVI

30  orf130-1.pep  YKNIAITFLLLHAAAE LWLPAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYYLLQ
    orf130ng-1   YKNIAIT-LLLHAAAE LWLPAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYYLLQ

35  orf130-1.pep  LFAAAGYLWTGA AKLQNL PASAPLHLITLGGMMGGVMMVWL TAGLWHS GFTKLDY PKLCR
    orf130ng-1   LFAAAGYLWTGA AKLQNL PASAPLHLITLGGMTGGVMMVWL TAGLWHS GFTKLDY PKLCR

40  orf130-1.pep  IAVPILFAAAVSRAFLMNVNPIFFITVPAILTAAVFVLYLFTFIPIFRANAFTDDPEX
    orf130ng-1   IAVSILFASAVSRAVLMNVNPIFFITVPEILTAAVFMLYLLTFVPIFRANAFTDDPEX

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 40 Example 101

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

```

1  ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
51  TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT
45  151 GCGGCGGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA
201 CGGCAATAGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGTA
351 CTGCTTGGA AAG..

```

50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

```

1  MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
51  GGESPPSLGD YEIPLSDGNS SVRANEYESA QSYFYRKIG KFEXCGLDWR
101 TRDGKPLIET FKQGGFDCLE K..

```

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

```

55  1  ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
    51  TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

```

-466-

5  
101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATT TTGGGATATT  
151 GCGGCGGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA  
201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT  
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA  
351 CTGCTTGGA AAGCAGGGGT TCGGCGCAA CCGTCTGTCC GAGCGCGTCC  
401 GATGGTAA

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

10  
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI  
51 GGESPPSLGD YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR  
101 TRDGKPLIET FKQGGFDCL E KQGLRRNGLS ERVRW\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N.*

15 *meningitidis*:

20  
orfl31a pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD  
orfl31a MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPPSLED  
25  
orfl31a pep YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCL E  
orfl31a YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK  
30  
orfl31a pep K  
orfl31a KQGLRRNGLSERVRWX  
130

The complete length ORF131a nucleotide sequence <SEQ ID 859> is:

35  
1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT  
51 TACGGTTGCA GGCTGCCGGT TGGCAGGTTG GTATGAGTGT TCGTCCCCTGT  
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATT TTGGGATATT  
151 GCGGCGGAGA GTCCTCCGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA  
201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT  
40  
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTGA  
351 TTGTTTGAAA AAGCAGGGGT TCGGCGCAA CCGTCTGTCC GAGCGCGTCC  
401 GATGGTAA

This encodes a protein having amino acid sequence <SEQ ID 860>:

45  
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI  
51 GGESPPSLED YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR  
101 TRDGKPLIET FKQEGFDCLK KQGLRRNGLS ERVRW\*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

50  
orfl31a pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPPSLED  
orfl31-1 MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD  
orfl31a pep YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK  
orfl31-1 YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCL E  
55  
orfl31a pep KQGLRRNGLSERVRWX  
|||||

orf131-1                    KQGLRRNGLSERVRWX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

5 *N.gonorrhoeae*:

	orf131.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD	60
		:         :           :	
	orf131.ng	MEIRVIKYTATAALFAFTVAGCRLAGWYECLSLGGWCKPRKPAAIDFWDIGGESPLSLED	60
10	orf131.pep	YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRD GKPLIETFKQGGFDCLE	120
		:                 :	
	orf131.ng	YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRD GKPLVERFKQEGFDCLE	120
15	orf131.pep	K	121
	orf131.ng	KQGLRRNGLSERVRW	134

A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

20

1	MEIRVIKYTA	TAALFAFTVA	GCRLAGWYEC	LSLSGWCKPR	KPAAIDFWDI
51	GGESPLSLED	YEIPLSDGNR	SVRANEYESA	QKSYFYRKIG	KFEACGLDWR
101	TRDGKPLVER	FKOEGFDCLE	KOGLRRNGLS	ERVVR*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

	1	ATGGAATTC	GGGTAATAAA	ATATACGGCA	ACGGCTCGGT	TGTTTGCATT
25	51	TACGGTTGCA	GGCTGCCGGC	TGCGGGGGTG	GTATGAGTGT	TCGTCCTTGT
	101	CCGGCTGGTG	TAAGCCGAGA	AAACCTGCCG	CCATCGATT	TTGGGATATT
	151	GGCGCGGAGA	GtcgcgtGTC	TTTAGAGGAC	TACGAGATCA	CGCTTTACAG
	201	CGGCAATCGT	TCCGTACGGG	CAAACGAATA	TGAATCCGCG	CAAAAATCTT
	251	ACTTTTATAG	GAAAATAGGG	AAGTTTGAAG	CCTGCGGGTT	GGATTGCGGT
30	301	ACGCGTGACG	GCAAACCTTT	GGTTGAGAGG	TTCAAACAGG	AAGGTTTCGA
	351	CTGTTTGGAA	AAGCAGGGGT	TGCGGCGCAA	CGGCTGTCC	GAGCGCGTCC
	401	GATGGTAA				

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

35

```

      1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SLSGWCKPR KPAIDFWDI
     51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
    101 TRDGKPLVER FKOEGFDCLE KOGLRNGLS ERVRW*
```

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

```

40      orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAADIFWDIGGESPLSLED
      orf131-1        MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSSLTGWCKPRKPAADIFWDIGGESPPSLGD

      orf131ng-1.pep YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
      orf131-1        YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE

45      orf131ng-1.pep KQGLRRNGLSERVRWX
      orf131-1        KQGLRRNGLSERVRWX

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be

50 useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 102**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
5  51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CCTGCCTTAT ATtTcCGGCC CGCAATGGCT GTCGGAAC
301 GTGCTGCACC ATCATTTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
10 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCCTCGCGC
401 CGGGCTTCCT TATtGGCGGC GTACC.GGAA AATtCGGCG TTTCCGCCCG
451 CTGCGCGCAA ACGCCGCGCC AAGACCCGAA CAGCCAATCG CCGTTTTTcG
501 TCATCGAAGC CGACGAATAC GACACCGCCT TTTTCGACAA ACGTTCTAAA
15 551 TtCGTGCAAT ACCGTCCGCG TACCGCCGTG TTGAACAATC TGGAATTCGA
601 CCACGCCGAC ATCTTTGCGC ACTTGGGCGC GATACAGACc CAGTTCCACT
651 ACCTCGTGCG TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGGACGG
701 CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGCTGGA
751 AAAATTCGGC ACGGAACACG GCTGGCA..

```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

20 1  MKHIHIIGIG GTFMGLLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51  YEGFDAAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSN
101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VXGKFRFRFP
151 PAANAAPRPE QPIAVERHRS RRIRHRLFRQ TFXIRALPSA YRRVEQSGIR
201 PRRHLCLRLGR DTDVPLPRA YRAVXRLNRL QRTAAKPARY FGQRLLDAGG
25 251 KIRHGTRLA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
30 101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CCTGCCTTAT ATTTCCGGCC CGCAATGGCT GTCGGAAC
301 GTGCTGCACC ATCATTTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
35 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
401 CGGGCTTCCT TATtGGCGGC GTACCGGAAA ATTTCCGGCGT TTCCGCCCGC
451 CTGCGCGCAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CCGTTTTTCGT
501 CATCGAAGCC GACGAATACG ACACCGCCTT TTTTCGACAA CCGTTCTAAAT
551 TCGTGCAATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC
601 CACGCCGACA TCTTTGCCGA CTTGGGCGCG ATACAGACCC AGTTCCACTA
40 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC
701 AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCCGTGGAA
751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGACGG
801 CTCGTTTCGAC GTGTTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAATGGG
45 851 ATTTGATGGG CAGGCACAAC CGCATGAACG CGCTCGCCGT CATTGCCGCC
901 GCGCGTCATG TCGGTGTCTGA TATTGAGACC GCCTGCGAAG CCTTGGGCGC
951 GTTTAAAAAC GTCAAACGCC GGATGGAAAT CAAAGGCACG GCAAACGGCA
1001 TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACCACG
1051 ATTCAAGGTT TGCGCCAACG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT
1101 CGAACCAGCT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC
50 1151 CTGTAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGGCGGCGTG
1201 CACTGGGACG TCGCGGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG
1301 TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
1351 GGAAAGCTGC TGGAAGCTTT GAGATAG

```

This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

```

1  MKHIHIIGIG GTFMGLLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51  YEGFDAAQLD EFKADVYVIG NVAKRGMDVV EAILNLGLPY ISGPQWLSN
60 101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
151 LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
201 HADIFADLGA IQTQFHYLVR TVPSEGLIVC NGRQQSLQDT LDKGCWTPVE
251 KFGTEHGWQA GEANADGSFD VLLDGKTAGR VKWDLMGRHN RMNALAVIAA
301 ARHVGVDIQT ACEALGAFKN VKRRMEIKGT ANGITVYDDF AHHPTAIETT

```



351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLKEA DQVFCYAGGV  
 401 DWDVAEALAP LGGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH  
 451 GKLEALR\*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of *E.coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

10 Orf132: 4 IHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLDEFK 63  
 IHI+GI GTFMGGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +  
 o457: 3 IHILGICGTFMGGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-Q 61  
 Orf132: 64 ADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTASML 123  
 D+ +IGN RG VEA+L +PY+SGPQWL + VL WVL VAGTHGKTTTA M  
 o457: 62 PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMA 121  
 15 Orf132: 124 AWWLEYAGLAPGFLIGGVXG 143  
 W+LE G PGF+IGGV G  
 o457: 122 TWILEQCGYKPGFVIGGVPG 141

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N. meningitidis*:

25 orf132.pep MKHHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLD  
 orf132a MKHHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLD  
 30 orf132.pep EFKADVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTAA  
 orf132a EFKADVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTAA  
 35 orf132.pep SMLAWVLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPRPEQPI-----AVFR  
 orf132a SMLAWVLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPRPEQPI-----AVFR  
 40 orf132.pep HRSRRIRHRLFRQTFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRL  
 orf132a KRSKFVHYRPRPTAVLNNLEFDHADIFADLGAIQTQFHHLVRTVPSEGLIVCNGRQQLQD

45 The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

1 ATGAAACACA TCCACATTAT CGGTATCGGC GGCACGTTTA TGGGTGGGAT  
 51 TGCCGCCATT GCCAAAGAAG CAGGGTTTGA ANTCAGCGGT TGCGATGCGA  
 101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG CATAGGCGTG  
 151 TATGAAGGCT TCGACACCGC GCAGTTGGAC GAATTTAAAG CCGACGTTTA  
 201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT  
 251 TGAACCGTGG GCTGCCTTAT ATTCCCGGCC CGCAATGGCT GGCTGAAAAC  
 301 NTGCTGCACC ATCATTTGNN ACTCGCGGTG GCGGNGACGC ACGGCAAAAC  
 351 GACCACCGCG TCGATGCTCG CGTGGGTTT GGAATATGCC GGAATTCGAC  
 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC  
 451 CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT  
 501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT  
 551 TCGTGCTATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC  
 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCACCA  
 651 CCTCGTGCGT ACCGTGCGGT CTGAAGGCCT CATCGTCTGC AACGGACGGC  
 701 AGCAAAGCCT GCAAGACACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAA  
 751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGATGG

	801	CTCGTTTCGAC	GTGTTGCTTG	ACGGCAAAAA	AGCCGGACAC	GTGCGTTGGA
	851	GTTTGATGGG	CGGACACAAC	CGCATGAACG	CGCTCGCNGT	CATCGCCGCC
	901	GCGCGTCATG	CCGGAGTNGA	CATTAGACG	GCCTCGGAAG	CCTTGAGCAG
5	951	GTTTAAAAAT	GTCAAAGCC	GCATGGAAT	CAAAAGCACG	GCAAACGGTA
	1001	TCACCGTTTA	CGACGACTTC	GCCCACCATC	CGACCGCTAT	CGAAACCACG
	1051	ATTCAAGGTT	TGCGCCAGCG	CGTCGGCGGC	GCGCGCATCC	TCGCCGCTCT
	1101	CGAACC CGGT	TCCAATACGA	TGAAGCTGGG	TACCATGAAA	GCCGCCCTGC
	1151	CCGCAAGCCT	CAAAGAAGCC	GACCAAGTGT	TCGTNTACGC	CGGCGCGCGC
	1201	GA CTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCGGCA	GGCTGCACGT
10	1251	CGGCAAAGAC	TTCTATGCCT	TCGTTGCCGA	AATCGTGAAA	AACGCCGAAG
	1301	CAGGCGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAAC TGC	TGGACGCTTT	GAGATAG		

15	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFEXSG	CDAKMYPPMS	TQLEALGIGV
	51	YEGFDTAQLD	EFKADVYVIG	NVAKRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	XLHHHWXLGV	AXTHGKTTTA	SMLAWVLEYA	GLAPGFXIGG	VPENFVSVAR
	151	LPQTPRQDPN	SQSPFFVIEA	DEYDTAFFDK	RSKFVHYHRP	TAVLNNLEFD
	201	HADIFADLGA	IQTQFHHLVR	TVPSEGLVC	NGRQVSLQDT	LDKGWCTPVE
20	251	KFGTEHGWA	GEANADGSFD	VLLDGKKAGH	VAWSLMGGHN	RMNALAVIAA
	301	ARHAGVDIQT	ACEALSTFKN	VKRRMEIKGT	ANGITVYDDF	AHHPATAIETT
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	AALPASLKEA	DQVFXYAGGA
	401	DWDVAEALAP	LGGRLHVGKD	FDAFVAEIVK	NAEAGDHILV	MSNGGFGGIH
	451	TKLLDALR*				

25	orf132a.pep	MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIGVYEGFDTAQLD
	orf132-1	MKHIHIIGIGGTFMGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAQLD
30	orf132a.pep	EFKADVYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAE NXLHHHWXLGVAXTHGKTTTA
	orf132-1	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA
35	orf132a.pep	SMLAWVLEYAGLAPGFXIGGVENFVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK
	orf132-1	SMLAWVLEYAGLAPGFLIGGVENFGVSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK
40	orf132a.pep	RSKFVHYRPERTAVLNNLEFDHADIFADLGAIQTQFHHLVRTVPSEGLIVCNGRQQSLQDT
	orf132-1	RSKFVHYRPERTAVLNNLEFDHADIFADLGAIQTQFHYLVRTVPSEGLIVCNGRQQSLQDT
45	orf132a.pep	LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKKAGHVAWSLMGGHNRMNALAVIAA
	orf132-1	LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRNMALAVIAA
50	orf132a.pep	ARHAGVDIQTACEALSTFKNVKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
	orf132-1	ARHVGVDIQTACEALGAFTKNVKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
55	orf132a.pep	ARILAVLEPRSNMTKLGTMKALPASLKEADQVFXYAGGADWDVAEALAPLGGRNLHV GKD
	orf132-1	ARILAVLEPRSNMTKLGTMKALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRNLV GKD
	orf132a.pep	FDAFVAEIVKNAEAGDHILVMSNGGFGGIHTKLLDALRX
	orf132-1	FDAFVAEIVKNAEVDHILVMSNGGFGGTHGKLLDALRX

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N. gonorrhoeae*:

```

60      orf132.pep  MKHIHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD      60
                  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      orf132ng   MKHIHIIGIGGTFMGGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAAOLE      60

```

-471-

	orf132.pep	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA	120
	orf132ng	EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA	120
5	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFRFRFPANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132ng	SMLAWVLEYAGLAPGFLIGGVPGKFRFRFPPTANAASRPEQQIAVFRHRSRRIRHRLFRQ	180
10	orf132.pep	TFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRLNRLQRTAAKPARY	240
	orf132ng	TLQIRALSPAYRRVEQSGIRPRRHLCRLGRDTPVPPRAHRTIRREHRLQRTAAKPARY	240
	orf132.pep	FGQRLLDAGGKIRHGTRLA	259
15	orf132ng	FGQRLLDAGGKIRHRLADW	261

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
20	51	HEGFDAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWVLGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPGKFRFRFP
	151	PTANAASRPE	QQIAVFRHRS	RRIRHRLFRQ	TLQIRALSPA	YRRVEQSGIR
	201	PRRHRLRGR	DTDPVPPRA	HRTIRRHRL	QRTAAKPARY	FGQRLLDAGG
	251	KIRHRLAD	W*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

25	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGCGGGAT
	51	TGCCGCCATT	GCCAAAGAAG	CCGGGTTCAA	AGTCAGCGGT	TGCGACGCGA
	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTA
	151	CACGAAGGCT	TCGATGCCGC	GCAGTTGGAA	GAATTTC AAG	CCGATATTTA
30	201	CGTCATCGGC	AATGTCGCCA	GGCGCGGGAT	GGATGTGGTC	GAGGCGATTT
	251	TGAACCGTGG	GCTGCCTTAT	ATTTCCGGCC	CGCAATGGCT	GGCTGAAAac
	301	GTGctgcacc	atcaTTGGgt	ACTCGGCGTG	GcagggaCGC	ACGGcaaAac
	351	gaccaCcGcg	tCCATGCTCG	CCTGGGTCTT	GGAATATGCC	GGACTGCGCG
	401	CGGGCTTCCT	CATCGGCGGt	gtaccggaAA	ATTTCGGCGT	TTCCGCCCGC
	451	CTACCGCAAA	CGCCGCGTCA	AGACCCGAAC	AGCAAATCGC	CGTTTTTCGT
35	501	CATCGAAGCC	GACGAATACG	ACACCGCCTT	TTTCGACAAA	CGCTCCAAAT
	551	TCGTGCATTA	TCGCCCGCGT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
	601	CACGCCGACA	TCTTCGCCGA	CTTGGGCGCG	ATACAGACCC	AGTTCACCA
	651	CCTCGTGC GC	ACCGTACCAT	CCGAAGGCCT	CATCGTCTGC	AACGGACAGC
	701	AGCAAAGCCT	GCAAGATACT	TTGGACAAAG	GCTGCTGGAC	GCCGTTGGAA
40	751	AAATTCGGCA	CCGGACACGG	CTGGCAGATT	GGTGAAGTCA	ATGCCGACGG
	801	CTCGTTCGAC	GTATTGCTTG	ACGGCAAAA	AGCCGGACAC	GTCGCTAGGG
	851	ATTTGATGGG	CGGACACAAC	CGCATGAACG	CGCTCGCCGT	CATCGCTGCC
	901	GCACGCCATG	CCGGAGTCGA	TGTTTCAGACG	GCCTGCGAAG	CCTTGGGTGC
	951	GTTTAAAAAC	GTCAAACGCC	GCATGGAAAT	CAAAGGCACG	GCAAACGGCA
45	1001	TCACCGTTTA	CGACGATTTT	GCCCACCACC	CGACCGCCAT	CGAAACCACG
	1051	ATTCAAGGTT	TGCGCCAACG	TGTCGGCGGC	GCGCGCATCC	TCGCCGTCTT
	1101	CGAGCCGCGT	TCCAACACCA	TGAAACTCGG	CACGATGAAG	TCCGCCCTGC
	1151	CCGCAAGCCT	CAAAGAAGCC	GACCAAGTGT	TCTGCTACGC	CGGCGGCGCG
	1201	GA CTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCTGCA	GGCTGCGCGT
50	1251	CGGTAAAGAT	TTCGATACCT	TCGTGCGCGA	AATTGTGAAA	AACGCCCAGG
	1301	CCGCGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAAC TGC	TGGACGCTTT	GAGATAG		

This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
55	51	HEGFDAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWVLGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPEHFVSAR
	151	LPQTPRODPN	SKSPFFVIEA	DEYDTAFDDK	RSKFVHYRPR	TAVLNNLEFD
	201	HADIFADLGA	IQTQFHHLVR	TVPSEGLIVC	NGQQQSLQDT	LDKGCWTPVE
	251	KFGTGHGWQI	GEVNADGSFD	VLLDGKKAGH	VAWDLMGGHN	RMNALAVIAA
60	301	ARHAGVDVQT	ACEALGAFKN	VKRRMEIKGT	ANGITVYDDF	AHHPTAIE TT
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	SALPASLKEA	DQVFCYAGGA
	401	DWDVAEALAP	LGCRLRVGKD	FDTFVAEIVK	NARTGDHILV	MSNGGFGGIH
	451	TKLLDALR*				

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

```

5  orf132ng-1.pep MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLE
   orf132-1      MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD

   orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTA
   orf132-1      EFKADVYVIGNVARRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA

10  orf132ng-1.pep SMLAWVLEYAGLAPGFLIGGVPEFVGSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDK
   orf132-1      SMLAWVLEYAGLAPGFLIGGVPEFVGSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK

   orf132ng-1.pep RSKFVHYRPRPTAVLNNLEFDHADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGQQQSLQDT
   orf132-1      RSKFVHYRPRPTAVLNNLEFDHADIFADLGAIQTQFHYLVVRTVPSEGLIVCNGRQQSLQDT

   orf132ng-1.pep LDKGCWTPVEKFGTGHGWQIGEVNADGSFDVLLDGKKAGHVAWDLMGHNRMNALAVIAA
   orf132-1      LDKGCWTPVEKFGTEHGWAQGEANADGSFDVLLDGKTAGRVKWDLMGRHNRMNALAVIAA

20  orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
   orf132-1      ARHVGVDIQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG

25  orf132ng-1.pep ARILAVLEPRSNMTMKLGTMKSLPASLKEADQVFCYAGGADWDVAEALAPLGCRLRVGKD
   orf132-1      ARILAVLEPRSNMTMKLGTMKSLPVSLKEADQVFCYAGGVDWDVAEALAPLGGRLNVGKD

30  orf132ng-1.pep FDTFVAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX
   orf132-1      FDAFVAEIVKNAEVGDHILVMSNGGFGGIHGKLLLEALRX

```

In addition, ORF132ng-1 is homologous to a hypothetical *E. coli* protein:

```

35  pir||S56459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)
   ORF_o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein
   in fbp-pmba intergenic region [Escherichia coli] Length = 457
   Score = 474 bits (1207), Expect = e-133
   Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)

40  Query: 22 KEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLEEFQADIYVIGNVARRGMDVVE 81
   ++ G +V+G DA +YPPMST LE GI + +G+DA+QLE Q D+ +IGN RG VE
   Sbjct: 21 RQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIIGNAMTRGNPCVE 79

45  Query: 82 AILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV 141
   A+L + +PY+SGPQWL + VL WVL VAGTHGKTTTA M W+LE G PGF+IGGV
   Sbjct: 80 AVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMATWILEQCGYKPGFVIGGV 139

50  Query: 142 PENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPRPTAVLNNLEFDH 201
   P NF VSA L +S FFVIEADEYD AFFDKRSKFVHY PRT +LNNLEFDH
   Sbjct: 140 PGNFEVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190

55  Query: 202 ADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGQQQSLQDTLDKGCWTPVEKFGTGHGWQIG 261
   ADIF DL AIQ QFHHLVR VP +G I+ +L+ T+ GCW+ E G WQ
   Sbjct: 191 ADIFDDLKAIQKQFHHLVRIVPGQGRIIWPENDINLKQTMAMGCWSEQELVGEQGHWQAK 250

60  Query: 262 EVNADGS-FDVLLDGKKAGHVAWDLMGHNRMNALAVIAAARHAGVDVQTACEALGAFKN 320
   ++ D S ++VLLDG+K G V W L+G HN N L IAAARH GV A ALG+F N
   Sbjct: 251 KLTTDASEWEVLLDGEKVGVEKWSLVGEHNMHNGLMAIAAARHVGVPADAANALGSFIN 310

65  Query: 321 VKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG-ARILAVLEPRSNMTMKLGT 379
   +RR+E++G ANG+TVYDDFAHHPTAI T+ LR +VGG ARI+AVLEPRSNMTMK+G
   Sbjct: 311 ARRLLELRGEANGVTVYDDFAHHPTAILATLAALRGKVGGTARIIVAVLEPRSNMTMKMGIC 370

   Query: 380 KSALPASLKEADQVF-CYAGGADWDVAEALAPLGCRLRVGKDFDTFVAEIVKNARTGDHI 438
   K L SL AD+VF W VAE D DT +VK A+ GDHI
   Sbjct: 371 KDDLAPSLGRADEVFLLQPAHIPWQVAEVAEACVQPAHWSGVDVTLADMVVKTAQPGDHI 430

   Query: 439 LVMSNGGFGGIHTKLLDAL 457

```

LVMSNGGFGGIH KLLD L  
 Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These  
 10 experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

### Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

```

1   . . CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA
15  51   CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC
    101  CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC
    151  ATTAGTGGCG ACTTCGGCGA TTATTTTCATG CCGTTCGCCA GCTATTCGCG
    201  CACACACCGT ATGCCCAACA TCCAAGAAAT GTATTTTTCC CAAATCGGCG
    251  ACTCCGGCGT TCACACCGCC TTAAAACCAG AGCGCGCAAA CACTTGGCAA
    301  TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAAAACAAG ATGATACATT
20  351  AGGATTAAAA CTGGTCGGCT ACCGCAGCCG CATCGACAAC TACATCCACA
    401  ACGTTTACGG GAAATGGTGG GATTTGAACG GGGATATTCC GAGCTGGGTC
    451  AGCAGCACCG GGCTTGCTTA CACCATCCAA CATCGCrATT TCaWAGACAA
    501  AGTGCATCAA nnnnnnnnnn nnnnnnnnnn nnnnTACGAT TATGGGCGTT
    551  TTTTCACCAA CCTTTCTTAC GCCTATCAAA AAAGCACGCA ACCGACCAAC
25  601  TTCAGCGATG CGAGCGAATC GCCCAACAAT GCGTCCAAAG AAGACCAACT
    651  CAAACAAGGT TATGGGTTGA GCAGGGTTTC CGCCCTGCCG CGAGATTACG
    701  GACGTTTGA AGTCGGTACG CGCTGGTTGG GCAACAAACT GACTTTGGGC
    751  GGCGCGATGC GCTATTTCGG CAAGAGCATC CGCGCGACGG CTGAAGAACG
    801  CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTTT CGGCAACTGG
30  851  GCAAGCGTTC CATCAAACAA ACCGAAACTC TTGCCCGCCA GCCTTTGATT
    901  TTWgATTTTa ACGCCGCTTA CGAGCCGAAG AAAAACCTTA TTTCCCGCGC
    951  CGAAGTCAAA AATCTGTTCG ACAGGCGTTA TATCGATCCG CTCGATGCGG
    1001 GCAATGATGC GGCAAC.GAG CGTTATTACA GCTCGTTCGA CCCGAAAGAC
    1051 AAGGACrrAG ACGTAACGTG TAATGCTGAT AAAACGTTGT GCaACGGCAA
35  1101 ATACGGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA
    1151 CCTTTTgAT GACGATGAGC TACAAGTTTT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

```

1   . . PGYYGSDDEF KRAFGENSPT XKKHCNRSCG IYEPVLKKYG KKRANNHSVS
40  51   ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ
    101  FGFXTYKKGL LKQDDTLGLK LVGYRSRIDN YIHNVYGKWW DLNGDIPSWV
    151  SSTGLAYTIQ HRFXDKVHQ XXXXXXXXXD YGRFFTNLSY AYQKSTQPTN
    201  FSDASESPNN ASKEDQLKQG YGLSRVSALP RDYGRLEVGT RWLGNKLTIG
    251  GAMRYFGKSI RATAEERYID GTNGGNTSNF RQLGKRSIKQ TETLARQPLI
    301  XDFNAAYEPK KNLIrFAEVK NLFDRRYIDP LDAGNDAAXE RYSSSFDPKD
45  351  KDXDVTcNAD KTLcNGKYGG TSKSVLTNFA RGRTFLMTMS YKF*
  
```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

```

1   GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
50  51   ACCGAAAGAC AAAAAAGTGT TTACCGATGC GCGTGCCGTA TCGACCCGTC
    101  AGGATATATT CAAATCCAGC GAAAACCTCG ACAACATCGT ACGCAGCATC
    151  CCCGGTGCGT TTACACAGCA AGATAAAAGC TCGGGCATTG TGTCTTTGAA
    201  TATTCGCGGC GACAGCGGGT TCGGGCGGGT CAATACGATG GTGGACGGCA
    251  TCACGCAGAC CTTTTATTCC ACTTCTACCG ATGCGGGCAG GGCAGGCGGT
  
```

5 301 TCATCTCAAT TCGGTGCATC TGTCGACAGC AATTTTATTG CCGGACTGGA  
 351 TGTCGTCAAA GGCAGCTTCA GCGGCTCGGC AGGCATCAAC AGCCTTGCCG  
 401 GTTCGGCGAA TCTGCGGACT TTAGGCGTGG ATGACGTCGT TCAGGGCAAT  
 451 AATACCTACG GCCTGCTGCT AAAAGGTCTG ACCGGCACCA ATTCAACCAA  
 501 AGGTAAATGCG ATGGCGGCGA TAGGTGCGCG CAAATGGCTG GAAAGCGGAG  
 551 CATCTGTGCG TGTCGCTTAC GGGCACAGCA GCGCGAGCGT GGCGCAAAAT  
 601 TACCGCGTGG GCGGCGGCGG GCAGCACATC GGAAATTTTG CCGCGGAATA  
 651 TTTGGAACGG CGCAAGCAGC GATATTTTGT ACAAGAGGGT GCTTTGAAAT  
 701 TCAATTCCGA CAGCGGAAAA TGGGAGCGGG ATTTACAAAG GCAACAGTGG  
 10 751 AAATACAAGC CGTATAAAAA TTACAACAAC CAAGAACTAC AaAAATACAT  
 801 CGAAGAGCAT GACAAAAGCT GGCGGAAAAA CCTg . CaCCG CAATACGACA  
 851 TTACCCCAT CGATCCGTCC AGCCTGAAGC AGCAGTCGGC AGGCAATCTG  
 901 TTTAAATTGG AATACGACGG CGTATTC AATATACACGG CGCAATTTCG  
 951 CGATTTAAC ACCAAAATCG GCAGCCGCAA AATCATCAAC CGCAATTATC  
 15 1001 AGTTCAATTA CGGTCTTCT TTTGAACCCGT ATACCAACCT CAATCTGACC  
 1051 GCAGCCTACA ATTGCGGCGAG GCAGAAATAT CCGAAAGGGT CGAAGTTTAC  
 1101 AGGCTGGGGG CTTTTAAAGG ATTTTGAAAC CTACAACAAC GCGAAAAATCC  
 1151 TCGACCTCAA CAACACCGCC ACCTTCCGGC TGCCCCGCGA AACCAGGTTG  
 20 1201 CAAACCACTT TGGGCTTCAA TTATTTCCAC AACGAATACG GCAAAAAACCG  
 1251 CTTTCTGAA GAATTGGGGC TGTTTTTCGA CGGTCTGAT CAGGACAACG  
 1301 GGCTTTATTC CTATTTGGGG CGTTTAAGG GCGATAAAGG GCTGCTGCCC  
 1351 CAAAATCAA CCATTGTCCA ACCGGCCGGC AGCCAATATT TCAACACGTT  
 1401 CTACTTCGAT GCGCGCTCA AAAAGACAT TTACCGCTTA AACTACGACA  
 25 1451 CCAATACCGT CGGCTACCGT TTCGGCGGGC AATATACGGG CTATTACGGC  
 1501 TCGGATGACG AATTTAAGCG GGCATTCGGA GAAAACTCGC CGACATACAA  
 1551 GAAACATTGC AACCAGGAGT GCGGGATTTA TGAACCCGTA TTGAAAAAAT  
 1601 ACGGCAAAAA GCGCGCCAAC AACCATTTCG TCAGCATTAG TCGCGACTTC  
 1651 GCGGATTATT TCATGCCGTT CGCCAGCTAT TCGCGCACAC ACCGTATGCC  
 30 1701 CAACATCCAA GAAATGTATT TTTCCCAAAT CCGCGACTCC GCGGTTACAC  
 1751 CCGCCTTAAA ACCAGAGCGC GCAAACTT GGCAATTTGG CTTCAATACC  
 1801 TATAAAAAAG GATTGTTAAA ACAAGATGAT ACATTAGGAT TAAAAGTGGT  
 1851 CCGCTACCGC AGCCGCATCG ACAACTACAT CCACAACGTT TACGGGAAAT  
 1901 GGTGGGATT GAACGGGGAT ATTCCGAGCT GGGTCAGCAG CACCGGGCTT  
 1951 GCCTACACCA TCCAACATCG CAATTTCAA GACAAAGTGC ACAAACACGG  
 35 2001 TTTTGAGTTG GAGCTGAATT ACGATTATGG GCGTTTTTTC ACCAACCTTT  
 2051 CTTACGCTTA TCAAAAAAGC ACGCAACCGA CCAACTTCAG CGATGCGAGC  
 2101 GAATCGCCCA ACAATGCGTC CAAAGAAGAC CAACTCAAAC AAGGTTATGG  
 2151 GTTGAGCAGG GTTCCGCCC TGCCGCGAGA TTACGGACGT TTGGAAGTCG  
 40 2201 GTACGCGCTG GTTGGGCAAC AACTGACTT TGGGCGGCGC GATGCGCTAT  
 2251 TTCGGCAAGA GCATCCGCGC GACGGCTGAA GAACGCTATA TCGACGGCAC  
 2301 CAACGGGGGA AATACCAGCA ATTTCCGGCA ACTGGCAAG CGTTCCATCA  
 2351 AACAAACCGA AACTCTTGCC CGCCAGCCTT TGATTTTGA TTTTACGCC  
 2401 GCTTACGAGC CGAAGAAAA CCTTATTTTC CGCGCGAAG TCAAAAATCT  
 45 2451 GTTCGACAGG CGTTATATCG ATCCGCTCGA TGCGGGCAAT GATGCGGCAA  
 2501 CGCAGCGTTA TTACAGCTCG TTCGACCCGA AAGACAAGGA CGAAGACGTA  
 2551 ACGTGTAATG CTGATAAAAC GTTGTGCAAC GGCAATACG CCGGCACAAG  
 2601 CAAAAGCGTA TTGACCAATT TTGCACGCGG ACGCACCTTT TTGATGACGA  
 2651 TGAGCTACAA GTTTTAA

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

50 1 EAQIQVLEDV HVKAKRVPKD KKVFTDARAV STRQDIFKSS ENLDNIVRSI  
 51 PGAFTQQDKS SGIVSLNIRG DSGFGRVNTM VDGITQTFYS TSTDAGRAGG  
 101 SSQFGASVDS NFIAGLDVVK GSFSGSAGIN SLAGSANLRT LGVDDVQGN  
 151 NTYGLLLKGL TGTNSTKGNA MAAIGARKWL ESGASVGVLY GHSRRSVAQN  
 201 YRVGGGGQHI GNFGAEYLER RKQRYFVQEG ALKFNSDSGK WERDLQRQQW  
 55 251 KYKPYKNYNN QELQYIEEH DKSWRENLXP QYDITPIDPS SLKQQSAGNL  
 301 FKLEYDGVFN KYTAQFRDLN TKIGSRKIIN RNYQFNYGLS INPYTNLNL  
 351 AAYNSGRQKY PKGSKFTGWG LLKDFETYNN AKILDNLNTA TFRLPRETEL  
 401 QTTLGFNHYFH NEYGKNRFFE ELGLFFDGPD QDNGLYSYLG RFGDKGLLP  
 451 KSTIVQPAG SQYFNTFFYD AALKKDIYRL NYSTNTVGYR FGGEYTGYYG  
 60 501 SDDEFKRAFG ENSPTYKKHC NRSCGIYEPV LKKYGKKRAN NHSVSIADP  
 551 GDYFMPFASY SRTHRMPIQ EMYFSQIGDS GVHTALKPER ANTWFQFENT  
 601 YKGLLLKQDD TLGLKLVGYS SRIDNYIHNV YGKWWDLNGD IPSWVSSTGL  
 651 AYTQHRNFK DKVHKHGFEL ELNYDYGRFF TNLSTAYQKS TQPTNFSAS  
 701 ESPNNASKED QLKQGYGLSR VSALPRDYGR LEVGTWRLGN KLTGGAMRY  
 751 FGKSIKIRATAE ERYIDGTNGG NTSNFRQLGK RSIKQETELA RQPLIFDFYA  
 801 AYEPKKNLIF RAEVKNLFD RYIDPLDAGN DAATQRYYS FDPKDKDEDV  
 851 TCNADKTLN KGYGTSKSV LTNFARGRTF LMTMSYKF\*

Computer analysis of this amino acid sequence gave the following results:

Orf133: 31 IYEPVLKKYGGKRRANNHVSISADFGDYFMPFASYSRTHRPNIQEMYFWSQIGDSGVHTA 90  
 I EP+L K G K+A NHS ++SA+ DYFMPF +YSRTHRPNIQEM+FSQ+ ++GV+TA  
 HI121: 563 INEPILHKSGHKKAFNHSATLSAELSDYFMPFFTYSRTHRPNIQEMFFSQVSNAGVNTA 622  
  
 Orf133: 91 LKPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWV 150  
 LKPE+++T+Q GF TYKKGL QDD LG+KLVGYRS I NYIHNVYG WW +P+W  
 HI121: 623 LKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNYIHNVYGVWW--RDGMPTWA 680  
  
 Orf133: 151 SSTGLAYTIQHRXFXDKVHXXXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSDDASESPNN 210  
 S G YTI H+ + V YD GRFF N+SYAYQ++ QPTN++DAS PNN  
 HI121: 681 ESNGFKYTIAHQNYKPIVKKSGVELEINYDMGRFFANVSYAYQRTNQPTNYADASPRPNN 740  
  
 Orf133: 211 ASKEDQLKQGYGLSRVSALPRDYGRLVGTWRWLGKNTLGLGAMRYFGKSIRATAEERYID 270  
 AS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTLG A RY+GKS RAT EE YI+  
 HI121: 741 ASQEDILKQGYGLSRVSMLEPKDYGRLELGTWFDQKLTGLAARYYGKSKRATIEEEYIN 800  
  
 Orf133: 271 GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDP 330  
 G+ + R+ ++K+TE + +QP+I D + +YEP K+LI +AEV+NL D+RY+DP  
 HI121: 801 GSR-FKKNLTLRRENYAVKKTEDIKKQPIILDLHVSYPKDLIIKAEVQNLLDKRYVDP 859  
  
 Orf133: 331 LDAGNDAAXERYSSFDPKDKDXDVTNADKTLNCGKYGGTSSKSVLTNFARGRTFLMTMS 390  
 LDAGNDAA +RYSS + + C D + C GG+ K+VL NFARGRT+++++  
 HI121: 860 LDAGNDAASQRYSSL-----NNSIECAQDSSAC---GGSDKTVLYNFGARGRTYILSLN 910  
  
 Orf133: 391 YKF 393  
 YKF  
 HI121: 911 YKF 913

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N. meningitidis*:

[illegible]

A partial ORF133a nucleotide sequence <SEQ ID 879> is:

1	AAAGACAAAA	AAGTGTTTAC	CGATGCGCGT	GCCGTATCGA	CCCGTCAGGA
51	TATATTCAAA	TCCANCGAAA	ACCTCGACAA	CATCGTACGC	ANCATCCCCG
101	GTGCGTTTAC	ACANCAANAT	AAAAGCTCGG	GCNATGTGTC	TTTGAATATT
151	CGCGCGACAC	CGGGGTTTCG	CGGGGTCAAT	ACNATGGTNG	ACGGCATCAC
201	NCANACCTTT	TATTGCACTT	CTACCGGATG	GGGACGGGCA	GGCGGTTTCA
251	CTCAATTTCG	TGCATCTGTC	GACAGCAATT	TTATNGCCGG	ACTGGATGTC
301	GTCAAAGGCA	GCTTCAGCGC	CTGCGCAGCG	ATCAACAGCC	TTGCCGGTTC
351	GCGGAATCTG	CGGACTTTAN	CGCTGGATGA	TGTCGTTTCA	GGCAATANTA
401	CNTACGCGCT	GCTGCTAAAA	GGTCTGACCG	GACCAATTCC	AACCAAGAGT
451	AATGCGATGG	CGGCGATAGG	TGCGCGCAAA	TGGCTGGAAA	GCGGAGCATC
501	TGTCGGTGTG	CTTTTACGGCG	ACAGCAGCGG	CAGCGTGGCG	CAAAAATTACC
551	GCGTGGGCGG	CGGCGGGCAG	CACATCGGAA	ATTTTGGCGC	GGAAATCTCG
601	GAACGACGCA	AGCAACGATA	TTTTTGAGCAA	GGAAGCGGGT	TGAAGTTTCAA
651	TTCCAACAGC	GGAAAATGGG	AGCGGGATTT	CCAAAAGTCG	TACTGGAAAA
701	CCAAGTGGTA	TCAAAAATAC	GATGCCCCCC	AAGAATCGCA	AAAATACATC
751	GAAGGTCATG	ATAAAAGCTG	CGGGGAAACC	CTGGCGGCCA	AATACGACAT
801	CACCCCATC	GATCGGTCCA	GCGTGAAGCN	CGAGTGGCGA	GGCAACCTGT
851	TTAAATTGGA	ATACGACGGC	GTATTCAATA	AATACACGGC	GCAATTTTCG
901	GATTTTAAACA	CCAAAATCGG	CAGCCGCAAA	ATCATCAACC	GCAATTTATCA
951	ATTTCAATTAC	GGTTTGTCTT	TGAACCCGTA	TACCAACCTC	AAATCTGACC
1001	CAGCCTACAA	TTCCGGGCAG	CAGAATATTC	GGAAGGGGTC	GGAAGTTTACA
1051	GGCTGGGGGC	TTTTNAAAGA	TTTTGAAACC	TACAACAACG	CAAAAATCCT
1101	CGACCTCANC	AACACCTCCA	CCTTCCGGCT	GCCCCGTGAA	ACCGAGTTGC
1151	AAACCACTTT	GGGCTTCAAT	TATTTCCACA	ACGAATTACG	CAAAAACCGC
1201	TTTCTCTAAG	AATTGGGGGT	GGTTTTCGAC	GGTCCGGATC	ANACACACGG
1251	GCTTTATTCC	TATTTGGGGC	GGTTTAAGGG	CGATAAAGGG	CTGCTGCCCC
1301	AAAAATCAAC	CATTGTCCAA	CCGGCCGGCA	GCCAATATTT	CAACACGTTT
1351	TACTTCGATC	CCGCGCTCAA	AAAAGACATT	TACCGCTTAA	ACTACAGCAT
1401	CAATACCGTC	GGCTACCGTT	TCGGCGGCNA	ATATACGGGC	TATTACNGCT
1451	CGGATGACGA	ATTTAAGCGG	GCATTTCGGG	AAAACCTCGC	GACATACANG
1501	AAACATTGCA	ACCAGAGCTG	CGGAATTTAT	GAACCCGTAT	TGAAAAAATA
1551	CGGCAAAAAG	CGCGCCAACA	ACCATTCCGT	CAGCATTAGT	GCGGACTTTC
1601	CGGATTAATT	CATGCGGTTT	CGCAGCTATT	CGCGCACACA	CCGTATGCCC
1651	AACATCCAAG	AAATGTATTT	TTCCCAAATC	GGCGACTCCG	GCGTTCACAC
1701	CGCCTTAAAA	CCAGAGCGCG	CAAAACATTT	GCAATTTGGC	TTCAATACCT
1751	ATAAAAAAGG	ATTGTTAAAA	CAAGATGATA	TATTAGGATT	AAAACCTGGT
1801	GGCTACCGCA	CGCGCATCGA	CNACTACATC	CACAACGTTT	ACGGGAATAG
1851	GTGGGATTTG	AACGGGAATA	TTCCGAGCTG	GGTCAGCAGC	ACCGGGCTTG
1901	CCTACACCAT	CCAACACCCG	AATTTCAAAG	ACAAAAGTGA	CAAAACCGGT
1951	TTTGAGTTGG	AGCTGAATTA	CGATTATNGG	CGTTTTTTTC	CCAACCTTTC
2001	TTACGGCTAT	CAAAAAGACA	CGCAACCGAC	CAACTTCAGC	GATGCGACGG
2051	AATCGCCCAA	CAATGCGTCC	AAAGAAGACC	AACTCAAACA	AGGTTATGGG
2101	TTGAGCAGGG	TTTCCGCCCT	GCCCGGAGAT	TACGGACGTT	TGGAAGTTCGG
2151	TACGCGCTGG	TTGGGCAACA	CAACTGACTT	GGCGGGCGCG	ATGCGCTATT
2201	TCGGCAAGAG	CATCCGCGCG	ACGGCTGAAG	AAACGTTATAT	CGACGNCACC
2251	AATGGGGNAN	NTACCAGCAA	TTTCCGGCAA	CTGGGCAAGC	GTTCCATCAN
2301	ACAAAACCGA	ACCCTTGCCC	GCCAGCCTTT	GATTTTTGAT	TTNTACGCCG
2351	CTTACGAGCC	GAAGAAAAAN	CTTATTTTTC	GCGCGGAAGT	CAAAAATCTG
2401	TTTCGACAGC	GTTATATFCG	TCCGCTTCGAT	GCGGGCAATG	ATGCGGCAAC
2451	GCAGCGTTAT	TACAGTTCGT	TCGACCCGAA	AGACAAGGAC	GAAGAAGTAA
2501	CGTGTAAATG	TGATACAACG	TTATGCAACG	GCAAAATACG	CGGCACAGC
2551	AAAAGCGTAT	TGACCAAAAT	TGCACGCGGA	CNCACCTTTT	TGATAACGAT
2601	GAGCTACAAG	TTTTAA			



This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

```

1   KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQX KSSGXVSLNI
5   51  RXDSGFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV
    101 VKGSFSGSAG INSLAGSANL RTLXVDDVVQ GNXTYGLLLK GLTGTNSTKG
    151 NAMAAGARK WLESGASVGV LYGHSRRSVA QNYRVGGGQQ HIGNFGAEYL
    201 ERRKQRYFEQ EGGLKFNSNS GKWERDFQKS YWKTWKYQKY DAPQELQKYI
    251 EGHDKSWREN LAPQYDITPI DPSSLKXQSA GNLFKLEYDG VFNKYTAQFR
    301 DLNTKIGSRK IINRNYQFNY GLSLNPYTNL NLTAAYNSGR QKYPKGSKFT
    351 GWGLXKDFET YNNAKILDIX NTSTFRLPRE TELQTTLGFN YFHNEYGKNR
10   401 FPEELGLFFD GPDNDGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF
    451 YFDAALKKDI YRLNYSTNTV GYRFGGXYTG YYXSDDEFKR AFGENSPTYX
    501 KHCNQSCGIY EPVLKKYGKK RANNHSVSIS ADFGDYFMPF ASYSRTHRMP
    551 NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGLLK QDDILGLKLV
    601 GYRSRIDXYI HNVYGKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG
15   651 FELELNIDYX RFTTNLSYAY QKSTQPTNFS DASESPNNAS KEDQLKQGYG
    701 LSRVSALPRD YGRLEVGTWR LGNKLTLGGA MRYFGKSIRA TAEERYIDXT
    751 NGXXTSNFRQ LCKRSIXQTE TLARQPLIFD XYAAEYPPKKX LIFRAEVKNL
    801 FDRRYIDPLD AGNDAATQRY YSSFDPKDKD EEVTCNDNDT LCNGKYGGTS
    851 KSVLTNFARG XTFLITMSYK F*

```

20 ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

```

                                10      20      30      40
orfl33a.pep                    KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGAFTXQXKS
25 orfl33-1                    EAQIQVLEDVHVKAKRVPKDKKVFTDARAVSTRQDIFKSSSENLDNIVRSIPGAFTQODKS
                                10      20      30      40      50      60

                                50      60      70      80      90      100
orfl33a.pep                    SGXVSLNIRXDSGFGRVNTMVDGITXTFYSTSTDAGRAGGSSQFGASVDSNFXAGLDVVK
30 orfl33-1                    SGIVSLNIRGDSGFGRVNTMVDGITQTFTYSTSTDAGRAGGSSQFGASVDSNFIAGLDVVK
                                70      80      90      100      110      120

                                110     120     130     140     150     160
orfl33a.pep                    GSFGSGSAGINSLAGSANLRTLXVDDVVQGNXTYGLLLKGLTGTNSTKGNAMAAIGARKWL
35 orfl33-1                    GSFGSGSAGINSLAGSANLRTLGVDDVVQGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWL
                                130     140     150     160     170     180

                                170     180     190     200     210     220
orfl33a.pep                    ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFEQEGGLKFNSNSGK
40 orfl33-1                    ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFVQEGALKFNSDSGK
                                190     200     210     220     230     240

                                230     240     250     260     270     280
orfl33a.pep                    WERDFQKS YWKTWKYQKYDAPQELQKYIEGHDKSWREN LAPQYDITPIDPSSLKXQ SAGN
50 orfl33-1                    WERDLQRQWKYKPYKNYNN-QELQKYIEEHDKSWREN LX PQYDITPIDPSSLKQ SAGN
                                250     260     270     280     290

                                290     300     310     320     330     340
orfl33a.pep                    LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
55 orfl33-1                    LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
                                300     310     320     330     340     350

                                350     360     370     380     390     400
orfl33a.pep                    YPKGSKFTGWGLXKDFETYNNAKILDIXNTSTFRLPRETELQTTLGFN YFHNEYGKNRFP
60 orfl33-1                    YPKGSKFTGWGLLKDFETYNNAKILDIXNTATFRLPRETELQTTLGFN YFHNEYGKNRFP
                                360     370     380     390     400     410

                                410     420     430     440     450     460
orfl33a.pep                    EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQ PAGSQYFNTFYFDAALKKDIYR
65 orfl33-1                    EELGLFFDGPQDNGLYSYLGRFKGDKGLLPQKSTIVQ PAGSQYFNTFYFDAALKKDIYR
                                420     430     440     450     460     470

```

-478-

		470	480	490	500	510	520
	orf133a.pep	LN	YSTNTVGYRFGGXYTGY	YXSDD	DEFKRAFGENSPTYK	HCNQSCGIYEPVL	KKYGKKRA
5	orf133-1	LN	YSTNTVGYRFGGEYTG	YYSDD	DEFKRAFGENSPTYK	HCNRS	CGIYEPVLKKYGKKRA
		480	490	500	510	520	530
	orf133a.pep						
10	orf133-1	NN	HSVSISADFGDYFMPFASYSR	THRMPNIQEMYFSQIGD	SGVHTALKPERANTWQ	FGFN	
		540	550	560	570	580	590
	orf133a.pep						
15	orf133-1	NN	HSVSISADFGDYFMPFASYSR	THRMPNIQEMYFSQIGD	SGVHTALKPERANTWQ	FGFN	
		600	610	620	630	640	
	orf133a.pep	TY	KKGLLKQDDILGLKLVGYRS	IDXYIHN	VYKWWDLNGNIPSWVS	STGLAYTIQHRNF	
	orf133-1	TY	KKGLLKQDDTLGLKLVGYRS	IDNYIHN	VYKWWDLNGDIPSWVS	STGLAYTIQHRNF	
		600	610	620	630	640	650
20	orf133a.pep						
	orf133-1	KD	KVHKHGFELNLDYGRFFTN	LSYAYQKSTQPTNFSDA	SESPNNASKEDQLKQ	GYGLS	
		660	670	680	690	700	710
25	orf133a.pep						
	orf133-1	RVS	ALPRDYGRLEVGT	RWLG	NKLT	LGGAMRYFGKSIRATAE	EERYIDXTNGXXTSNFRQLG
		720	730	740	750	760	770
30	orf133a.pep						
	orf133-1	RVS	ALPRDYGRLEVGT	RWLG	NKLT	LGGAMRYFGKSIRATAE	EERYIDGTNGGNTSNFRQLG
		720	730	740	750	760	770
35	orf133a.pep						
	orf133-1	KRS	IXQTETLARQPLIFDX	YAAEY	PKKXLI	FRAEVKNLFDRRYID	PLDAGNDAATQRYYS
		780	790	800	810	820	830
40	orf133a.pep						
	orf133-1	SF	DPKDKDEEVT	CNDNTLCNGKYGGT	SKSVLT	NFARGXTFLITMSYKFX	
		840	850	860	870	880	

Homology with a predicted ORF from *N.gonorrhoeae*ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N.*45 *gonorrhoeae*:

	orf133.pep	PGYYGSDDEFKRAFGENSPTXKKHCNRS CGI	31
	orf133ng	FYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAFGENSPAYKEHCDPSCGL	560
50	orf133.pep	YEPVLKKYGKKRANNHSHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKKYGKKRANNHSHSVSISADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL	620
55	orf133.pep	KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS	151
	orf133ng	KPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVG	680
60	orf133.pep	STGLAYTIQHRXFXDKVHQXXXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFS DASESPNNA	211
	orf133ng	STGLAYTIRHRNFKDKVHKHGFELNLDYDYGRFFTNLSYAYQKSTQPTNFS DASESPNNA	740
	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGT RWLG NKLT LGGAMRYFGKSIRATAEERYIDG	271
	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGT RWLG NKLT LGGAMRYFGKSIRATAEERYIDG	800
65	orf133.pep	TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFD RRYIDPL	331
	orf133ng	TNGGNTSNVRQLGKRSIKQTETLARQPLIFD FYAAYEPKKNLIFRAEVKNLFD RRYIDPL	860

orf133.pep    DAGNDAAXERYSSFDPKDKDXDVTCTNADKTLNKGKYGGSKSVLTNFARGRTFLMTMSY    391  
 |||||:|||||||  
 orf133ng    DAGNDAATQRYSSFDPKDKDEDVTCTNADKTLNKGKYGGSKSVLTNFARGRTFLMTMSY    920  
 5  
 orf133.pep    KF    393  
 ||  
 orf133ng    KF    922

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a  
 10 protein having amino acid sequence <SEQ ID 882>:

1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVAKAKRV  
 51 PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN  
 101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGSSSQFGAS VDSNFIAGLD  
 151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK  
 201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY  
 251 LERRKQQYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYK YEDPQELQKY  
 301 IEEHDKSWRE NLAPQYDITP IDPSGLKQOS AGNLLMLEYD GVFNKYTAQF  
 351 RDLNTRIGSR KIINRNYQFN YGLSLNPTYN LNLTAAYNSG RQKYPKGAKF  
 401 TGWGLLKDFE TYNNAKILD L NNTATFRLPR ETELQTTLGF NYFHNEYGKN  
 20 451 RFPEELGLFF DGPDQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT  
 501 FYFDAALKKD IYRLNYSTNA INYRFGGEYT GYGSSENEFK RAFGENSPAY  
 551 KEHCDPSCGL YEPVLKKGK KRANNHSVSI SADFGDYFMP FAGYSRTHRM  
 601 PNIQEMYFSQ IGDGSHVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL  
 651 VGYRSRIDNY IHNVYKWWDD LNGDIPSWVG STGLAYTIRH RNFKDKVHKH  
 25 701 GFELELNIDY GRFFTNLSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY  
 751 GLSRVSALPR DYGRLEVGT R WLGKLTLLGG AMRYFGKSIR ATAERYIDG  
 801 TNGGNTSNVR QLGKRSIKQT ETLARQPLIF DFYAAVEPKK NLI FRAEVKN  
 851 LFDRLYIDPL DAGNDAATQR YSSFDPKDK DEDVTCTNADK TLCNKGKYGGS  
 901 SKSVLTNFAR GRTFLMTMSY KF\*

30 A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

1 ATGAGATCTT CTTTCCGGTT GAAGCCGATT TGTTTTATC TTATGGGTGT  
 51 TATGCTATAT CATCATAGTT ATGCCGAAGA TGCAGGGCGC GCGGGCAGCG  
 101 AGGCGCAGAT ACAGGTTTGT GAAGATGTGC ACGTCAAGGC GAAGCGCGTA  
 151 CCGAAAGACA AAAAAGTGTT TACCGATGCG CGTGCCGTAT CGACCCGTca  
 201 gGATGTGTTC AAATCCGGCG AAAACCTCGA CAACATCGTA CGCAGCATAC  
 251 CCGGTGCGTT TACACAGCAA GATAAAAGCT CGGGCATTGT GTCTTTGAAT  
 301 ATTCGCGGCG ACAGCGGGTT CGGGCGGGTC AATACGATGG TGGACGGCAT  
 351 CACGCAGACC TTTTATTCTGA CTTCTACCGA TCGGGCAGG GCAGGCGGTT  
 401 CATCTCAATT CGGTGATCTT GTCGACAGCA ATTTTATTGC CGGACTGGAT  
 40 451 GTCGTCAAAG GCAGCTTCAG CGGCTCGGCA GGCATCAACA GCCTTGCCGG  
 501 TTCGCGCAAT CTGCGGACTT TAGGCGTGGA TGACGTCTGT CAGGGCAATA  
 551 ATACCTACGG CCTGCTGCTA AAAGGTCTGA CCGGCACCAA TTCAACCAAA  
 601 GGTAATGCGA TGGCGCGGAT AGGTGCGCGC AAATGGCTGG AAAGCGGAGC  
 651 GTCTGTCTGG GTGCTTTACG GGCACAGCAG GCGCGGCGTG GCGCAAAATT  
 45 701 ACCGCGTGGG CGGCGGCGGG CAGCACATCG GAAATTTTGG TGAAGAAAT  
 751 CTGGAACGGC GCAAACAGCA ATATTTTGTG CAAGAGGGTG GTTTGAAATT  
 801 CAATGCCGGC AGCGGAAAAT GGAACGGGA TTTGCAAAG CAATACTGGA  
 851 AAACAAAGTG GTATAAAAAA TACGAAGACC CCAAGAAGT GCAAAAATAC  
 901 ATCGAAGAGC ATGATAAAAG CTGGCGGGAA AACCTGGCGC CGCAATACGA  
 50 951 CATCACCCCC ATCGATCCGT CCGGCCTGAA GCAGCAGTCG GCAGGCAATC  
 1001 TGTTTAAATT GGAATACGAC GGCGTATTCA ATAAATACAC GGCGCAATTT  
 1051 CGCGATTTAA ACACCAGAA CTGGCAGCCG AAAATCATCA ACCGCAATTA  
 1101 TCAATTCAAT TACGGTTTGT CTTTGAACCC GTATACCAAC CTCAATCTGA  
 1151 CCGCAGCCTA CAATTGGGGC AGGCAGAAAT ATCCGAAAGG GGCGAAGTTT  
 55 1201 ACAGGCTGGG GGCTTTTAAA AGATTTTGAA ACCTACAACA ACGCGAAAAT  
 1251 CCTCGACCTC AACAACACCG CCACCTTCCG GCTGCCCGCG GAAACCGAGT  
 1301 TGCAAAACCAC TTTGGGCTTC AATTATTTCC ACAACGAATA CGGCAAAAC  
 1351 GCCTTTCCTG AAGAATTGGG GCTGTTTTTC GACGGTCCTG ATCAGGACAA  
 1401 CGGGCTTTAT TCCTATTGGG GGCGGTTTAA GGGCGATAAA GGGCTGTTCG  
 60 1451 CTCAAAAATC AACCATTTGTC CAACCGGCCG GCAGCCAATA TTTCAACACG  
 1501 TTCTACTTCG ATGCCGCGCT CAAAAAGAC ATTTACCGCT TAAACTACAG  
 1551 ACCCAATGCA ATCAACTACC GTTTCGGCGG CGAATATACG GGCTATTACG  
 1601 GCTCGGAAAA CGAATTTAAG CGGGCATTTC GAGAAAACCT GCCGGCATA  
 1651 AAGGAACATT GCGACCCGAG CTGCGGGCTT TATGAACCCG TATTGAAAAA  
 65 1701 ATACGGCAAA AAGCGCGCCA ACAACCATTC GGTACGATT AGTCGGGACT  
 1751 TCGGCGATTA TTTCATGCCG TTCGCCGGCT ATTCGCGCAC ACACCGTATG

-480-

1801 CCCAACATCC AAGAAATGTA TTTTCCCAA ATCGGCGACT CCGGCGTTCA  
 1851 CACCGCCTTA AAACCAGAGC GCGCAAACAC TTGGCAATTT GGCTTCAATA  
 1901 CCTATAAAAA AGGATTGTTA AAACAAGATG ATATATTAGG ATTGAAACTG  
 1951 GTCGGCTACC GCAGCCGCAT TGACAACTAC ATCCACAACG TTACCGGGAA  
 2001 ATGGTGGGAT TTGAACGGGG ATATCCGAG CTGGGTCGGC AGCACC GGCG  
 2051 TTGCCTACAC CATCCGACAC CGCAATTTCA AAGACAAAGT GCACAAACAC  
 2101 GGTTTTGAGC TGGAGCTGAA TTACGATTAT GGGCGTTTTT TCACCAACCT  
 2151 TTCTTACGCC TATCAAAAA GCACGCAACC GACCAATTC AGCGATGCGA  
 2201 GCGAATCGCC CAACAATGCC tccaaAGAAG ACCAACTCAA ACAAGGTTAT  
 2251 GGGCTGAGCA GGGTTTCCGC CCTGCCGCGA GATTACGGAC GTTGGAAAGT  
 2301 CCGTACGCGC TGGTTGGGCA ACAAACTGAC TTTGGGCGGC GCGAtgcGT  
 2351 ATTTCCGGCAA GAGCATCCGC GCGACGGCTG AAGAACGCTA TATCGACGGC  
 2401 ACCAACGGGG GAAATACCAG CAATGTCCGG CAACTGGGCA AGCGTTCCAT  
 2451 CAAACAAACC GAAACCCTTG CCCGACAGCC TTTGATTTT GATTTTACG  
 2501 CCGTTACGCA GCCGAAGAAA AACCTTATTT TCCGCGCGCA AGTCAAAAAC  
 2551 CTGTTGACGA GCGGTTATAT CGATCCGCTC GATGCGGGCA ATGATGCGGC  
 2601 AACGCAGCGT TATTACAGCT CGTTCGACCC GAAAGACAAG GACGAAGACG  
 2651 TAACGTGTAA TGCTGATAAA ACGTTGTGCA ACGGCAAATA CGGCGGCACA  
 2701 AGCAAAAGCG TATTGACCAA TTTTCGCACG GGACGCACCT TCTTGATGAC  
 2751 GATGAGCTAC AAGTTTTAA

This corresponds to the amino acid sequence <SEQ ID 884; ORF133ng-1>:

1 MRSSFRLKEI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVKAARV  
 51 PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN  
 101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGGSSQFGAS VDSNFIAGLD  
 151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK  
 201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY  
 251 LERRKQQYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYKK YEDPQELQKY  
 301 TEEHDKSWRE NLAPQYDITP IDPSGLKQQS AGNLFKLEYD GFVNKYTAQF  
 351 RDLNTRIGSR KIINRNYQFN YGLSLNPHYTN LNLTAAYNSG RQKYPKGAKE  
 401 TGGWLLKDFE TYNNAKILD LNTATFRLPR ETELQTTLGF NYFHNEYGNK  
 451 RFPEELGLFF DGPQDQNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT  
 501 FYFDAALKKD IYRLNYSTNA INYRFGGEYT GYYGSENEFK RAFGENSPAY  
 551 KEHCDPSCGL YEPVLKKYK KKRNNHVSIV SADFGDYFMP FAGYSRTHRM  
 601 PNIQEMYFSQ IGD SGVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL  
 651 VGYRSRIDNY IHN VYGKWWD LNGDIPSWVG STGLAYTIRH RNFKDKVHKH  
 701 GFELELNIDY GRFTTNLSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY  
 751 GLSRVSALPR DYGRLEVGR WLGNKLTGG AMRYFGKSIR ATAEERYIDG  
 801 TNGNTSNVR QLGRSISIKT ETLARQPLIF DFYAAYPEKK NLIFRAEVKN  
 851 LFDRRYIDPL DAGNDAATQR YYSSFDPKDK DEDVTCNADK TLCNGKYGGT  
 901 SKSVLTNFA GRTEFLMTMSY KF\*

ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:

10 20 30 40 50 60  
 orf133ng-1.pep SFRLKPICFYLMGVMLYHHSYAEDAGRAGSEAQIQVLEDDVHVKAARV  
 45 orf133-1 EAQIQVLEDDVHVKAARV  
 10 20 30  
 70 80 90 100 110 120  
 orf133ng-1.pep STRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS  
 50 orf133-1 STRQDIFKSSSENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS  
 40 50 60 70 80 90  
 130 140 150 160 170 180  
 55 orf133ng-1.pep TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN  
 orf133-1 TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN  
 100 110 120 130 140 150  
 190 200 210 220 230 240  
 60 orf133ng-1.pep NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRGVAQNYRVGGGGQHI  
 orf133-1 NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRSVAQNYRVGGGGQHI  
 160 170 180 190 200 210  
 250 260 270 280 290 300  
 65 orf133ng-1.pep GNFGEEYLERRKQQYFVQEGGLKFNAGSGKWERDLQRQYWKTKWYKKYEDPQELQKYIEE

-481-

```

|||||
orfl133-1  GNFGAEYLERRKQRYFVOEGALKFNSDSGKWERDLQRQQWKYPKYNYN-QELQKYIEE
              220      230      240      250      260

5
orfl133ng-1.pep  HDKSWRENLAPOYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
orfl133-1        HDKSWRENLPQYDITPIDPSSLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
              270      280      290      300      310      320

10
orfl133ng-1.pep  NRNYQFNYSLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
orfl133-1        NRNYQFNYSLSLNPYTNLNLTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDNLNT
              330      340      350      360      370      380

15
orfl133ng-1.pep  ATFRLPRETELQTTLGFNYPFHNEYGKNRFEELGLFFDGPQDNGLYSYLGRFKGDKGLL
orfl133-1        ATFRLPRETELQTTLGFNYPFHNEYGKNRFEELGLFFDGPQDNGLYSYLGRFKGDKGLL
              390      400      410      420      430      440

20
orfl133ng-1.pep  PQKSTIVQPAGSQYFNTFYFDAALKKDIYRLNYSNTNAINYRFGGEYTGYYGSENEFKRAF
orfl133-1        PQKSTIVQPAGSQYFNTFYFDAALKKDIYRLNYSNTNTVGYRFGGEYTGYYGSDDEFKRAF
              450      460      470      480      490      500

25
orfl133ng-1.pep  GENSPAYKEHCDPSCGLYEPVLKKYGKKRANNHVSISADFGDYFMPFAGYSRTHRMPNI
orfl133-1        GENSPYKHKCNRSYEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNI
              510      520      530      540      550      560

30
orfl133ng-1.pep  QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHN
orfl133-1        QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNYIHN
              570      580      590      600      610      620

35
orfl133ng-1.pep  VYGKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
orfl133-1        VYGKWWDLNGDIPSWVSTGLAYTIQHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
              630      640      650      660      670      680

40
orfl133ng-1.pep  STQPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMR
orfl133-1        STQPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMR
              690      700      710      720      730      740

45
orfl133ng-1.pep  YFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAYEPKKNLI
orfl133-1        YFGKSIRATAEERYIDGTNGGNTSNFROLGKRSIKQTETLARQPLIFDYAAYEPKKNLI
              750      760      770      780      790      800

50
orfl133ng-1.pep  FRAEVKNLFDRRYIDPLDAGNDAATQRYSSFDPKDKDEDVTCNADKTLGNGKYGGTSKS
orfl133-1        FRAEVKNLFDRRYIDPLDAGNDAATQRYSSFDPKDKDEDVTCNADKTLGNGKYGGTSKS
              810      820      830      840      850      860

55
orfl133ng-1.pep  VLTNFARGRTFLMTMSYKFX
orfl133-1        VLTNFARGRTFLMTMSYKFX
              870      880

60
65
70

```

In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H. influenzae*:

sp|P45114|YC17\_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR HI1217 PRECURSOR  
>gi|1075372|pir||G64110 transferrin binding protein 1 precursor (tbpl) homolog -  
Haemophilus influenzae (strain Rd KW20) >gi|1574147 (U32801) transferrin binding  
protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913  
Score = 930 bits (2377), Expect = 0.0  
Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

5  
Query: 38 QVLEDVHVHAKRVPKDKKVFDDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV 97  
+ L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V  
10  
Sbjct: 29 ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKSGGVV 88

Query: 98 SLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGSSQFGASVDSNFIAGLDVVKGSFS 157  
S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS  
15  
Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGSSQFGAAIDPNFIAGVDVKNKSNFS 148

Query: 158 GSAGINSLAGSANLRTLGVDDVVQXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217  
G++GIN+LAGSAN RTLGV+DV+ M RKWL++G  
Sbjct: 149 GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFMTMAAGRKWLNDGG 208

20  
Query: 218 SVGVLYGHSRRGVAQNYRVGGGGQHIGNFGEELYERRKQQYFVQEGGLKFNAGSGKWERD 277  
VGV+YG+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D  
Sbjct: 209 YVGVVYGYSQREVSQDYRI-GGGERLASLGQDILAKEKEAYF-RNAGYILNP-EGQWTPD 265

25  
Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303  
L +++W +Y KK +D ++LQK IEE  
Sbjct: 266 LSKKHWSCNKPQDYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKNGDIEE 325

30  
Query: 304 HDKSWRENAPQYDITPIDPSGLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363  
DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI  
Sbjct: 326 TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFEYGDHQNGLAQLRLDNLKIGSRKIE 384

35  
Query: 364 NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT 423  
NRNYQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+  
Sbjct: 385 NRNYQVNYFNNNNSYLDLNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNS 444

40  
Query: 424 ATFRLPRETELQTTLGFNYFHNEYGKNRFPEELGLFFDGPDQDNGLYSY--LGRFGKDKG 481  
TF LP+E +L+TTLGFNYF NEY KNRPEEL LF++ D GLYS+ GR+ G K  
Sbjct: 445 HTFLLPKEIDLKTTLGFNYFTNEYSKNRFPEELSLFYNDASHDQGLYSHSKRGRYSGTKS 504

45  
Query: 482 LLPQKSTIVQFAGSQYFNTFYFDAALKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKR 541  
LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY  
Sbjct: 505 LLPQRSVILQPSGKQKFTVYFDALSKGIYHLNYSVNFTHYAFNGEYVGY----- 555

50  
Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKGKRRANNHVSISADFGDYFMPFAGYSRTHRMP 601  
EN+ + + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP  
Sbjct: 556 ---ENTAGQQ-----INEPILHKS GHKAFNHSATLSAELSDYFMPFFTYSRTHRMP 604

55  
Query: 602 NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYI 661  
NIQEM+FSQ+ ++GV+TALKPE+++T+Q GFNTYKKG L QDD+LG+KLVGYRS I NYI  
Sbjct: 605 NIQEMFFSQVSNAAGVNTALKPEQSDTYQLGFNTYKKGLETDVLDVGLVGYRSFIKNYI 664

60  
Query: 662 HNVYGKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELNLDYGRFFTNLSYAY 721  
HNVYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY  
Sbjct: 665 HNVYGVWW--RDGMPTWAESNGFKYTIHQNYPKIVKKS GVELEINYDMGRFFANVSYAY 722

65  
Query: 722 QKSTQPTNFS DASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGNKLTGGA 781  
Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTLG A  
Sbjct: 723 QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMLEPKDYGRLELGTRWFDQKLTGLA 782

70  
Query: 782 MRYFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAYEPKKN 841  
RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+  
Sbjct: 783 ARYYGKSKRATIEEYINGSR-FKNTLRRENYAVKKTEDIKKQPIILDHVSYEPIKD 841

Query: 842 LIFRAEVKNLFDRRYIDPLDAGNDAATQRYSSFPDKDKDEDVTCNADKTLNGKYGTS 901  
LI +AEV+NL D+RY+DPLDAGNDAA+QRYSS + + C D + C GG+  
Sbjct: 842 LIIKAEVQNLLDKRYVDPLDAGNDAASQRYSSSL-----NNSIECAQDSSAC----GGSD 892

Query: 902 KSVLTNFAFGRTFLMTMSYKF 922  
K+VL NFARGRT+++++YKF  
Sbjct: 893 KTVLYNFARGRTYILSLNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

```

1   ATGAACCTGA  TTTCACGTTA  CATCATCCGT  CAAATGGCGG  TTATGGCGGT
51  TTACGCGCTC  CTTGCCTTCC  TCGCTTTGTA  CAGCTTTTTT  GAAATCCTGT
101 ACGAAACCGG  CAACCTCGGC  AAAGGCAGTT  ACGGCATATG  GGAAATGCTG
151 GGCTACACCG  CCCTCAAAAT  GCGCGCCGCG  GCCTACGAAC  TGATTCCCCT
201 CGCCGTCCCT  ATCGGCGGAC  TGGTCTCCCT  CAGCCAGCTT  GCCGCCGGCA
251 GCGAACTGAC  CGTCATCAAA  GCCAGCGGCA  TGAGCACCAA  AAAGCTGCTG
301 TTGATTCTGT  CGCAGTTCGG  TTTTATTTTT  GCTATTGCCA  CCGTCGCGCT
351 CGGCGAATGG  GTTGCGCCCA  CACTGAGCCA  AAAAGCCGAA  AACATCAAAG
15  401 CCGCCGCCAT  CAACGGCAAA  ATCAGCACCG  GCAATACCGG  CCTTTGGCTG
451 AAAGAAAAAA  ACAGCGTGAT  CAATGTGCGC  GAAATGTTGC  CCGACCAT..

```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```

1   MNLISRYIIR  QMAVMAVYAL  LAFLALYSFF  EILYETGNLG  KGSYGIWEML
51  GYTALKMPAR  AYELIPLAVL  IGGLVLSLSQL  AAGSELTVIK  ASGMSTKKLL
20  101 LILSQFGFIF  AIATVALGEW  VAPTLSQLAE  NIKAAAINGK  ISTGNTGLWL
151 KEKNSVINVR  EMLPDH...

```

Further work revealed further partial nucleotide sequence <SEQ ID 887>:

```

1   ATGAACCTGA  TTTCACGTTA  CATCATCCGT  CAAATGGCGG  TTATGGCGGT
25  51  TTACGCGCTC  CTTGCCTTCC  TCGCTTTGTA  CAGCTTTTTT  GAAATCCTGT
101 ACGAAACCGG  CAACCTCGGC  AAAGGCAGTT  ACGGCATATG  GGAAATGCTG
151 gGCTACACCG  CCCTCAAAAT  GCGCGCCGCG  GCCTACGAAC  TGATTCCCCT
201 CGCCGTCCCT  ATCGGCGGAC  TGGTCTCCCT  CAGCCAGCTT  GCCGCCGGCA
251 GCGAACTGAC  CGTCATCAAA  GCCAGCGGCA  TGAGCACCAA  AAAGCTGCTG
301 TTGATTCTGT  CGCAGTTCGG  TTTTATTTTT  GCTATTGCCA  CCGTCGCGCT
30  351 CGGCGAATGG  GTTGCGCCCA  CACTGAGCCA  AAAAGCCGAA  AACATCAAAG
401 CCGCCGCCAT  CAACGGCAAA  ATCAGCACCG  GCAATACCGG  CCTTTGGCTG
451 AAAGAAAAAA  ACAGCrTKAT  CAATGTGCGC  GAAATGTTGC  CCGACCATAC
501 GCTTTTGGGC  ATCAAAATTT  GGGCGCGCAA  CGATAAAAC  GAATTGGCAG
551 AGGCAGTGGA  AGCCGATTCC  GCCGTTTGA  ACAGCGACGG  CAGTTGGCAG
35  601 TTGAAAAACA  TCCGCCGCAG  CACGCTTGGC  GAAGACAAAG  TCGAGGTCTC
651 TATTGCGGCT  GAAGAAAAC  GGCCGATTTC  CGTCAAACGC  AACCTGATGG
701 ACGTATTGCT  CGTCAAACCC  GACCAAATGT  CCGTCGGCGA  ACTGACCACC
751 TACATCCGCC  ACCTCCAAAA  CAACAGCCAA  AACACCCGAA  TCTACGCCAT
801 CGCATGGTGG  CGCAAATTGG  TTTACCCGCG  CGCAGCCTGG  GTGATGGCGC
40  851 TCGTCGCCTT  TGCCTTTACC  CCGCAAACCA  CCCGCCACGG  CAATATGGGC
901 TTAAAACTCT  TCGCGGGCAT  CTGTsTCGGA  TTGCTGTTCC  ACCTTGCCCG
951 ACGGCTCTTT  GGGTTTACCA  GCCAACTCGG...

```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```

1   MNLISRYIIR  QMAVMAVYAL  LAFLALYSFF  EILYETGNLG  KGSYGIWEML
45  51  GYTALKMPAR  AYELIPLAVL  IGGLVLSLSQL  AAGSELTVIK  ASGMSTKKLL
101 LILSQFGFIF  AIATVALGEW  VAPTLSQLAE  NIKAAAINGK  ISTGNTGLWL
151 KEKNSXINVR  EMLPDHTLLG  IKIWARNDKN  ELAEAVEADS  AVLNSDGSWQ
201 LKNIRRSTLG  EDKVEVSIAA  EENWPISVKR  NLMDVLLVKP  DQMSVGELTT
50  251 YIRHLQNSQ  NTRIYAIAWW  RKLVPAAAW  VMALVAFAT  PQTRHGNMG
301 LKLFGGICXG  LLFHLAGRLF  GFSTSQL...

```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.*

*meningitidis*:

5	orf112.pep	10	20	30	40	50	60
	orf112a	10	20	30	40	50	60
10	orf112.pep	70	80	90	100	110	120
	orf112a	70	80	90	100	110	120
15	orf112.pep	130	140	150	160		
	orf112a	130	140	150	160		
20	orf112a	130	140	150	160	170	180
	orf112a	190	200	210	220	230	240

The ORF112a nucleotide sequence <SEQ ID 889> is:

25	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGT	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCCTTC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
30	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGNTG
	151	GGNTACACCG	CCCTCAAAAT	GNCCGCCCGC	GCCTACGAAC	TGATGCCCTT
35	201	CGCCGTCCTT	ATCGGCGGAC	TGGTCTCTNT	CAGCCAGCTT	GCCGCCGGCA
	251	GCGAACTGAN	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAG	AAAGCTGCTG
40	301	TTGATTCTGT	CGCAGTTCGG	TTTTATTTT	GCTATTGCCA	CCGTGCGGCT
	351	CGGCGAATGG	GTTGCGCCCA	CACTGAGCCA	AAAAGCCGAA	AACATCAAAG
45	401	CCGCGGCCAT	CAACGGCAAA	ATCAGTACCG	GCAATACCGG	CCTTTGGCTG
	451	AAAGAAAAAA	ACAGCATTAT	CAATGTGCGC	GAAATGTTGC	CCGACCATAC
50	501	CCTGCTGGGC	ATTAAATCT	GGGCCCGCAA	CGATAAAAC	GAAGTGGCAG
	551	AGGCAGTGGG	AGCCGATTCC	GCCGTTTGA	ACAGCGACCG	CAGTTGGCAG
55	601	TTGAAAAACA	TCCGCGCGAG	CACGCTTGGC	GAAGACAAAG	TGAGGTCTC
	651	TATTGCGGCT	GAAGAAAAANT	GGCCGATTTC	CGTCAAACGC	AACCTGATGG
60	701	ACGTATTGCT	CGTCAAACCC	GACCAAATGT	CCGTGCGCGA	ACTGACCACC
	751	TACATCCGCC	ACCTCCAAAN	NNACAGCCAA	AACACCCGAA	TCTACGCCAT
65	801	CGCATGGTGG	CGCAAATTGG	TTTACCCCGC	CGCAGCCTGG	GTGATGGCGC
	851	TCCGTCGCTT	TGCCCTTACC	CCGCAAAACCA	CCCGCCACGG	CAATATGGGC
70	901	TTAAAANTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTCC	ACCTTGCCGG
	951	NCGGCTCTTC	NGGTTTACCA	GCCAACTCTA	CGGCATCCCG	CCCTTCTCTG
75	1001	NCGGCGCACT	ACCTACCATA	GCCTTCGCCT	TGCTCGCCGT	TTGGCTGATA
	1051	CGCAAACAGG	AAAAACGCTA	A		

This encodes a protein having the amino acid sequence <SEQ ID 890>:

50	1	MNLISRYIIR	QMAVMAYYAL	LAFLALYSFF	EILYETGNLG	KGSYGIWEMX
	51	GYTALKMXAR	AYELMPLAVL	IGGLVXSXQL	AAGSELXVIK	ASGMSTKKLL
55	101	LILSQFGFIF	AIATVALGEW	VAPTLSQKAE	NIKAAAINGK	ISTGNTGLWL
	151	KEKNSIINVR	EMLPDHTLLG	IKIWARNDKN	ELAEAVEADS	AVLNSDGSWQ
60	201	LKNIRIRSTLG	EDKVEVSIAA	EEXWPISVKR	NLMDVLLVKP	QMSVGEELT
	251	YIRHLQXXSQ	NTRIYIAIWW	RKLVPAAAW	VMALVAFAPT	PQTRHGNMG
65	301	LKXFGGICLG	LLFHLAARLF	XFTSQLYGIP	PFLXGALPTI	AFALLAVWLI
	351	RKQEK*				

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

60	orf112a.pep	MNLISRYIIRQMAVMAYYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR
	orf112-1	MNLISRYIIRQMAVMAYYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR
	orf112a.pep	AYELMPLAVLIGGLVXSXQLAAGSELXVIKASGMSTKKLLILSQFGFIFAIATVALGEW



10

15

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from *N. gonorrhoeae*:

25

30

40

45

50

55

The complete length ORF112ng nucleotide sequence <SEQ ID 891> is:

35

60

12-1	AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLLLILSQFGFIFAIATVALGEW
12a.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN
12-1	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSXINVREMLPDHTLLGIKIWARNDKN
12a.pep	ELAAEVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEEXWPISVKRNLMDVLLVKP
12-1	ELAAEVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEENWPISVKRNLMDVLLVKP
12a.pep	DQMSVGELTTYIRHLQXXSQNTRIYAIAWWRKLVYPAAAWVMALVAFATPQTTRHGNMG
12-1	DQMSVGELTTYIRHLQNNSQNTRIYAIAWWRKLVYPAAAWVMALVAFATPQTTRHGNMG
12a.pep	LKXFGGICLGLLFHLAGRLFXFTS QLYGIPPLFXGALPTIAFALLAVWLIRKQEKRX
12-1	LKLFGGICXGLLFHLAGRLFGFTSOL

orf112.pep	MNLISRYIIIRQMAMVAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	60
orf112ng		
orf112.pep	MNLISRYIIIRQMAMVAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	60
orf112ng		
orf112.pep	AYELIPLAVLIGGLVLSLSQLAAGSELTVIKASGMSTKKLLILSQFGFIFAIATVALGEW	120
orf112ng		
orf112.pep	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLILSQFGFIFAAI AVALGEW	120
orf112ng		
orf112.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH	166
orf112ng		
orf112.pep	VAPTLSOKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN	180
orf112ng		

1	ATGAACCTGA	TTTCACGTTA	CATCATCCGC	CAAATGGCGG	TTATGGCGGT
51	TTACGCGCTC	CTTGCCCTCC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
101	ACGAAACCGG	CAACCTCGGC	AAAGCGAGTT	ACGGCATATG	GGAAATGCTG
151	GGCTACACCG	CCCTCAAAAT	GCCGCGCCGC	GCCTACGAAC	TCATGCCCCT
201	CGCGCTCCTC	ATCGCGGCAC	TGGCTTCTCT	CAGGCAGCTT	GCCGCGCGCA
251	GCGAACTGGC	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
301	TTGATTCTGT	CTCAGTTCGG	TTTTATTTTT	GCTATTGCCG	CCGTGCGCGT
351	CGCGGAATGG	GTTCGCGCCA	CGCTGAGCCA	AAAGCGCGAA	AACATCAAag
401	cCGCGCCCA	taacggyCAAA	ATACGACA	gcAATACCGG	CCCTTggcTG
451	AAAGAAAAAA	ccAGCATTAT	CAATGTGcGc	GGAATGTTGC	CCGACCATAC
501	GCTTTTGGGC	ATCAAAATTT	GGGCGCGCAA	CGATAAAAAA	GAATTGGCAG
551	AGGCAGTGGC	AGCCGATTCC	GCCGTTTTGA	ACAGCGACGG	GACCTGGCAG
601	TTGAAAAACA	TCCGCGCGAC	CATCATGGGT	ACAGACAAAA	TCGAAACATC
651	cgCGCGCGCC	GAAGAAACTT	gGCCGATTGC	CGTCAGACGC	AACCTGATGG
701	ACGTATTGCT	CGTCAAGCCC	GACCAAATGT	CCGTGCGCGA	GCTGACCACC
751	TACATCCGCG	ACCTCAAAAA	CAACAGCCAA	AACACCCAAA	TCTACGCCAT
801	CGCATGGTGG	CGTAAACTCG	TTTACCCCGT	CGCGCATATG	GTCAATGGCGC
851	TCGTTGCGCTT	CGCCTTTACG	CCGCAAAACA	CGCGCCACGG	CAATATGGGC
901	TTAAAACTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTCC	ACCTTGCCCG
951	CAGGCTCTTC	GGGTTTACCA	GCCAACTCTA	CGGCACCCCA	CCCTTCCTCG
1001	CGGCGCGCAT	CGCTACCATA	GCCTTCGCCT	TGCTCGCTGT	TTGGCTGATA
1051	CGCAAAACAGG	AAAAACGTTG	A		

This encodes a protein having amino acid sequence <SEQ ID 892>:

1	<u>MNLISRYIIR</u>	<u>QMAVMAYVAL</u>	<u>LAFLALYSFF</u>	<u>EILYETGNLG</u>	<u>KSGSYGIWEML</u>
51	<u>GYTALKMPAR</u>	<u>AYELMPLAVL</u>	<u>TGGLASLSQL</u>	<u>AAGSELAVIK</u>	<u>ASGSMYTKLL</u>
101	<u>LILSQFGFIF</u>	<u>AIAAVALGEW</u>	<u>VAPTLTSQAK</u>	<u>NIKAAAIKNG</u>	<u>ISTGNTGLWL</u>
151	<u>KEKTSIINVR</u>	<u>GMLPDPHTLLG</u>	<u>IKIWARDKN</u>	<u>ELAEAVEADS</u>	<u>AVLNSDGSWQ</u>
201	<u>LKNIRRSIMG</u>	<u>TDKIEYISAAA</u>	<u>EETWPTAVRR</u>	<u>NLMDVLLVKP</u>	<u>DQMSVGELTT</u>
251	<u>YIRHLQNNQS</u>	<u>NTQIYAIAWW</u>	<u>RKLVPYVAAR</u>	<u>VMALVAFAT</u>	<u>PQTRRHGNMG</u>
301	<u>LKLFGGTCLG</u>	<u>LLFHLAGRLF</u>	<u>GFTSLQYGTP</u>	<u>PFLAGALPTI</u>	<u>AFALLAWVLI</u>

351 RKQEK\*  
\*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
5	orf112ng	MNLISRYII	RQMAVM	AVYALLA	FLALYS	SFFEILY	ETGNLGKGSYGIWEM
	orf112-1	MNLISRYII	RQMAVM	AVYALLA	FLALYS	SFFEILY	ETGNLGKGSYGIWEM
		10	20	30	40	50	60
10	orf112ng	70	80	90	100	110	120
		AYELMPLA	VLIGGLA	SLSQLA	AGSELAV	IKASGMST	KKLLLSQFGFI
	orf112-1	AYELIPLA	VLIGGLV	SLSQLA	AGSELTV	IKASGMST	KKLLLSQFGFI
		70	80	90	100	110	120
15	orf112ng	130	140	150	160	170	180
		VAPTLSQK	AENIKAA	AINKGIS	TGNTGL	WLKEKTS	IINVRGMLPD
	orf112-1	VAPTLSQK	AENIKAA	AINKGIS	TGNTGL	WLKEKNS	XINVREMLPD
		130	140	150	160	170	180
20	orf112ng	190	200	210	220	230	240
		ELAEAVEA	DSAVLNS	DGSQLKN	IRRSIM	GTDKIETS	SAAAEETWPI
	orf112-1	ELAEAVEA	DSAVLNS	DGSQLKN	IRRSIM	GTDKIETS	SAAAEETWPI
		190	200	210	220	230	240
25	orf112ng	250	260	270	280	290	300
		DQMSVGEL	TTYIRHL	QNNNSQ	NTQIYA	IAWWRKL	VYPVAAWVM
	orf112-1	DQMSVGEL	TTYIRHL	QNNNSQ	NTQIYA	IAWWRKL	VYPVAAWVM
		250	260	270	280	290	300
30	orf112ng	310	320	330	340	350	
		LKLFGGIC	LGLLFHL	AGRLFG	FTSQLY	GTTPFFLA	GALPTIAF
	orf112-1	LKLFGGIC	LGLLFHL	AGRLFG	FTSQL		
		310	320				

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTCTGG	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-CCAGCGGTAGCCTAATT	
ORF 2	Forward	GCGGATCCCATATG-TTTGATTTCTGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GACGGCATAACGGCG	
ORF 2-1	Forward	GCGGATCCCATATG-TTTGATTTCTGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGATTTACGGACGCGCA	
ORF 4	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTGGCTGCGCCTTC	
ORF 5	Forward	GGAATTCCATATGGCCATGG-TGGAAGGCGCACAAAC	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-ATGGAAGGCGCACAAAC	
	Reverse	CCCGCTCGAG-GACTGTGCAAAAACGG	
ORF 6	Forward	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGCGCCGAACACTTTC	
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTGTGTTCC	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-TTTCAAAATATATTTGCGGA	
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGCAGGCTTTGGCGC	
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGGAAA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTCAGAGTTTTCGGG	
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TAATGGGAAACCTTGTTTT	
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GGAAACGACTTCGCC	
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTTCCGCGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGGGTGTGTGATAATAAG	
ORF 15	Forward	GGAATTCCATATGGCCATGG-GCGGGACACTGACAG	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-TGCGGGACACTGACAGG	
	Reverse	CCCGCTCGAG-AGGTTGGCCTTGCTATG	
ORF 17	Forward	GGAATTCCATATGGCCATGG -TTGCCGGCCTGTTTCG	NdeI-NcoI

<b>ORF 18</b>	Forward	CGGGATCC-ATTGCCGCGCTGTTTCG	BamHI
	Reverse	CCCGCTCGAG-AAGCAGGTGTACAGC	XhoI
<b>ORF 19</b>	Forward	GCGGATCCCATATG-ATTTGCTGCATTTGGAT	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCTTCCAATTTCTGAAAGC	XhoI
<b>ORF 19</b>	Forward	GGAATTCCATATGGCCATGG -TCGCCAGTGT TTTTACC	NdeI-NcoI
	Forward	CGGGATCC-TTCGCCAGTGT TTTTACCG	BamHI
	Reverse	CCCGCTCGAG-GGTGT TTTTGAAGCTGCC	XhoI
<b>ORF 20</b>	Forward	GGAATTCCATATGGCCATGG -TCGGCGCGGGTATG	NdeI-NcoI
	Forward	CGGGATCC-TTCGGCGCGGGTATG	BamHI
	Reverse	CCCGCTCGAG-CGGCGAGCGAGAGCA	XhoI
<b>ORF 22</b>	Forward	GGAATTCCATATGGCCATGG-TGATTAAAATCAAAAAAGGTCT	NdeI-NcoI
	Forward	CGGGATCC-ATGATTAAAATCAAAAAAGGTCTAAACC	BamHI
	Reverse	CCCGCTCGAG-ATTATGATAGCGGCC	XhoI
<b>ORF 23</b>	Forward	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTAAACCGATAGGTAAACG	XhoI
<b>ORF 24</b>	Forward	GGAATTCCATATGGCCAIGG -TGATGCCGGAATGGTG	NdeI-NcoI
	Forward	CGGGATCC-ATGATGCCGGAATGGTG	BamHI
	Reverse	CCCGCTCGAG-TGTCAGCGTGCGCA	XhoI
<b>ORF 25</b>	Forward	GCGGATCCCATATG-TATCGCAAATGATTGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCGATGGAATAGCCG	XhoI
<b>ORF 26</b>	Forward	GCGGATCCCATATG -CAGCTGATCGACTATTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACATCGCGCGT TTT	XhoI
<b>ORF 27</b>	Forward	GGAATTCCATATGGCCATGG-AGACCTATTCTGT TTA	NdeI-NcoI
	Forward	CGGGATCC- CAGACCTATTCTGT TTTATTTTAATC	BamHI
	Reverse	CCCGCTCGAG-GGGTTCGATTAAATAACCAT	XhoI
<b>ORF 28</b>	Forward	GGAATTCCATATGGCCATGG-ACGGCTGTACGTTGATGT	NdeI-NcoI
	Forward	CGGGATCC-AACGGCTGTACGTTGATG	BamHI
	Reverse	CCCGCTCGAG-TTTGTCAGAGGAATTCGCG	XhoI
<b>ORF 29</b>	Forward	GCGGATCCCATATG -AACGGTTTGGATGCCCCG	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-AACGGTTTGGATGCCCCG	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTGTCTAAGTTCTCTGATATG	XhoI
<b>ORF 32</b>	Forward	CGCGGATCCCATATG-AATACTCCTCCTTTTG	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGTATTTT T T T T GATGCTTTG	XhoI
<b>ORF 33</b>	Forward	GCGGATCCCATATG -ATTGATAGGGATCGTATG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGATCTTTCAAACGGCC	XhoI

<b>ORF 35</b>	Forward Forward Reverse	GCGGATCCCATATG-TTCAGAGCTCAGCTT CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT CCCGCTCGAG-AAACAGCCATTTGAGCGA	BamHI-NdeI BamHI-NheI XhoI
<b>ORF 37</b>	Forward Reverse	GCGGATCCCATATG-GATGACGTATCGGATTTT CCCGCTCGAG-ATAGCCCGCTTTCAGG	BamHI-NdeI XhoI
<b>ORF 58</b>	Forward Reverse	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT CCCGCTCGAG-AGCATTGTCCAAGGGGAC	BamHI-NheI XhoI
<b>ORF 65</b>	Forward  Forward Reverse	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG  CGGGATCC-TTGCTGTATCTGAATCAAGG CCCGCTCGAG-CCGCATCGGCAGACA	NdeI-NcoI BamHI XhoI
<b>ORF 66</b>	Forward Reverse	GCGGATCCCATATG-TACGCATTTACGCGG CCCGCTCGAG-TGGATTTTGCAGAGATGG	BamHI-NdeI XhoI
<b>ORF 72</b>	Forward Reverse	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA CCCGCTCGAG-GCCTGAGACCTTTGCAA	BamHI-NdeI XhoI
<b>ORF 73</b>	Forward Reverse	GCGGATCCCATATG-AGATTTTTTCGGTATCGG CCCGCTCGAG-TTCATCTTTTTTCATGTTG	BamHI-NdeI XhoI
<b>ORF 75</b>	Forward Reverse	GCGGATCCCATATG- TCTGTCTTTCAAACGGC CCCGCTCGAG-TTTGTTTTTGCAAGACAG	BamHI-NdeI XhoI
<b>ORF 76</b>	Forward Reverse	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC CGGGATCC-TTACGGTTTGACACCGTT	NheI-NdeI BamHI
<b>ORF 79</b>	Forward Reverse	CGCGGATCCCATATG-GTTTCCGCCGCCG CCCGCTCGAG-GTGCTGATGCGCTTCG	BamHI-NdeI XhoI
<b>ORF 83</b>	Forward Reverse	GCGGATCCCATATG-AAAACCTGCTGCTGC CCCGCTCGAG-GCCGCCTTTGCGGC	BamHI-NdeI XhoI
<b>ORF 84</b>	Forward Reverse	GCGGATCCCATATG-GCAGAGATCTGTTTG CCCGCTCGAG-GTTTGCCGATCCGACCA	BamHI-NdeI XhoI
<b>ORF 85</b>	Forward Reverse	CGCGGATCCCATATG- GCGGTTTGGGGCGGA CCCGCTCGAG-TCGGCGCGGCGGGC	BamHI-NdeI XhoI
<b>ORF 89</b>	Forward Forward Reverse	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA CGGGATCC-GCCATACCTTCTTATCAGAG CCCGCTCGAG-TTTTTTGCGATTAGAAAAAGC	NdeI-NcoI BamHI XhoI
<b>ORF 97</b>	Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC	BamHI-NdeI

<b>ORF 98</b>	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTG	XhoI
	Forward Reverse	GCGGATCCCATATG-ACGGTAACTGCGG CCCGCTCGAG-TTGTGTTCGGGCAAATC	BamHI-NdeI XhoI
<b>ORF 100</b>	Forward Reverse	GCGGATCCCATATG-TCGGGCATTTACACCG CCCGCTCGAG-ACGGGTTTCGGCGGAA	BamHI-NdeI XhoI
<b>ORF 101</b>	Forward Reverse	GCGGATCCCATATG-ATTTATCAAAGAAACCTC CCCGCTCGAG-TTTCCGCCTTTCAATGT	BamHI-NdeI XhoI
<b>ORF 102</b>	Forward Reverse	GCGGATCCCATATG-GCAGGGCTGTTTTACC CCCGCTCGAG-AAACGGTTTGAACACGAC	BamHI-NdeI XhoI
<b>ORF 103</b>	Forward Reverse	GCGGATCCCATATG-AACCACGACATCAC CCCGCTCGAG-CAGCCACAGGACGGC	BamHI-NdeI XhoI
<b>ORF 104</b>	Forward Reverse	GCGGATCCCATATG-ACGTGGGGAACGC CCCGCTCGAG-GCGGCGTTTGAACGGC	BamHI-NdeI XhoI
<b>ORF 105</b>	Forward Reverse	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC CCCGCTCGAG-TAAACGAATGCCGTCCAG	BamHI-NdeI XhoI
<b>ORF 106</b>	Forward Reverse	GCGGATCCCATATG-AGGATAACCGACGGCG CCCGCTCGAG-TTTGTTCCCGATGATGTT	BamHI-NdeI XhoI
<b>ORF 109</b>	Forward Reverse	GCGGATCCCATATG-GAAGATTTATATATAATACTCG CCCGCTCGAG-ATCAGCTTCGAACCGAAG	BamHI-NdeI XhoI
<b>ORF110</b>	Forward Reverse	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC AAACTGCAG-GGAAAACCACATCCGCACTCTGCC	EcoRI PstI
<b>ORF111</b>	Forward Reverse	AAAGAATTC-GCACCGCAAAAGGCAAAACCGCA AAACTGCAG-TCTGCGCGT TTTGGGCAGGGTGG	EcoRI PstI
<b>ORF113</b>	Forward Reverse	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTTCAACCG AAACTGCAG-TTACGAATGCCTGCTTGCTCGACCGTACTG	EcoRI PstI
<b>ORF115</b>	Forward Reverse	AAAGAATTC-TTGCTTGTCGAAACAGAAAAAGACGG AAAAAGTCGAC-CTATTTTTTAGGGGCTTTTGTGTTTGAAGCCTGCC	EcoRI SalI
<b>ORF119</b>	Forward Reverse	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG AAACTGCAG-TTATGAAACAGGCGCAGGGCGGTTTTGCC	EcoRI PstI
<b>ORF120</b>	Forward Reverse	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG AAACTGCAG-CGGTTTGGCTGCCTGGCCGTTGAT	EcoRI PstI
<b>ORF121</b>	Forward Reverse	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTTCGC AAACTGCAG-TCATCCGCCACCCACCTCGGCCATCCATC	EcoRI PstI

<b>ORF122</b>	Forward Reverse	AAAAAAGTCGAC-ATGTC TTACCGCGCAAGCAGTTC TCC AAACTGCAG-TCAGGAACACAAACGATGACGAATATCCGTATC	Sall PstI
<b>ORF125</b>	Forward Reverse	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT AAACTGCAG-CGCCGTTTCAAGACGAAAAAGTCG	EcoRI PstI
<b>ORF126</b>	Forward Reverse	AAAGAATTC-GCGGAAACGGTCGAAG AAACTGCAG-TTAATCTTGTCTCCGATATAC	EcoRI PstI
<b>ORF127</b>	Forward Reverse	AAAGAATTC-ATGACTGATAATCGGGGGTTTACG AAAAAAGTCGAC-CTTAAGTAAGTGCAGTCCTTATC	EcoRI Sall
<b>ORF128</b>	Forward Reverse	AAAGAATTC-ATGCAAGCTGTCCGCTACAGGCC AAACTGCAG-CTATTGCAATGCGCCGCCGCGGAATGTTGAGCAGGCG	EcoRI PstI
<b>ORF129</b>	Forward Reverse	AAAGAATTC-ATGGATTTTCGTTTTGACATTATTTACGAATACCG AAACTGCAG-TTATTTTTTGATGAAATTTTGGGGCGG	EcoRI PstI
<b>ORF130</b>	Forward Reverse	AAAGAATTC-GCAGTACTTGCCATTCTCGGTGCG AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT	EcoRI PstI
<b>ORF 131</b>	Forward Reverse	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT CCCGCTCGAG-CCAGCGGACGCGTTC	BamHI-NdeI XhoI
<b>ORF 132</b>	Forward Reverse	GCGGATCCCATATG-AAAGAAGCGGGGTTTG CCCGCTCGAG-CCAATCTGCCAGCCGT	BamHI-NdeI XhoI
<b>ORF 133</b>	Forward Reverse	CGCGGATCCCATATG-GAAGATGCAGGGCGCG CCCGCTCGAG-AAACTTGTAGCTCATCGT	BamHI-NdeI XhoI
<b>ORF 134</b>	Forward Reverse	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG CCCGCTCGAG-ATCCTGTGCCAATGCG	BamHI-NdeI XhoI
<b>ORF 135</b>	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT CCCGCTCGAG-AAATACCGCTGAGGATG	BamHI-NdeI XhoI
<b>ORF 136</b>	Forward Reverse	CGCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC CCCGCTCGAG-TTCCGAATATTTGGAACTTTT	BamHI-NheI XhoI
<b>ORF 137</b>	Forward Reverse	CGCGGATCCCATATG-GGCACGGCGGGAAATA CCCGCTCGAG-ATAACGGTATGCCGCC	BamHI-NdeI XhoI
<b>ORF 138</b>	Forward Reverse	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC CCCGCTCGAG-CGGCGTTTTTATAGCGG	BamHI-NdeI XhoI
<b>ORF 139</b>	Forward Reverse	GCGGATCCCATATG-GCTTTTTTGCGGTAATG CCCGCTCGAG-TAACGTTTCCGTGCGTTT	BamHI-NdeI XhoI

<b>ORF 140</b>	Forward Reverse	GCGGATCCCATATG-TTGCCACAGGCAGC CCCGCTCGAG-GACGATGGCAAACAGC	BamHI-NdeI XhoI
<b>ORF 141</b>	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAGCAGTCT CCCGCTCGAG-ATCTGTTGTTTTTAAATATT	BamHI-NdeI XhoI
<b>ORF 142</b>	Forward Reverse	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG CCCGCTCGAG-AAACGTATAGCCTACCT	BamHI-NdeI XhoI
<b>ORF 143</b>	Forward Reverse	GCGGATCCCATATG-GATACCGCTTTGAACCT CCCGCTCGAG-AATGGCTTCCGCAATATG	BamHI-NdeI XhoI
<b>ORF 144</b>	Forward Reverse	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC CCCGCTCGAG-AGATTGTTGTTGTTTTTCG	BamHI-NdeI XhoI
<b>ORF 147</b>	Forward Reverse	GCGGATCCCATATG-TCTGTCTTTCAAACGGC CCCGCTCGAG-TTGTTTTTGCAAGACAG	BamHI-NdeI XhoI

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SalI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SalI* site was used in the reverse primer.



TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

**CLAIMS**

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.
2. A nucleic acid molecule which encodes a protein according to claim 1.
- 5 3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 10 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 15 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 20 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 25 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
- 30 5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..

7. An antibody which binds to a protein according to any one of claims 4 to 6.

8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.

9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

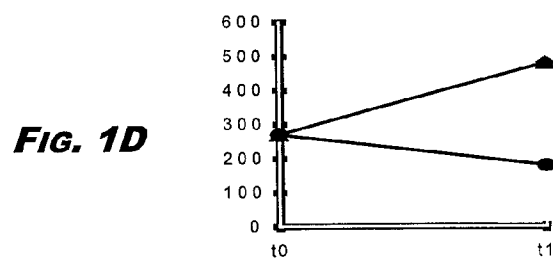
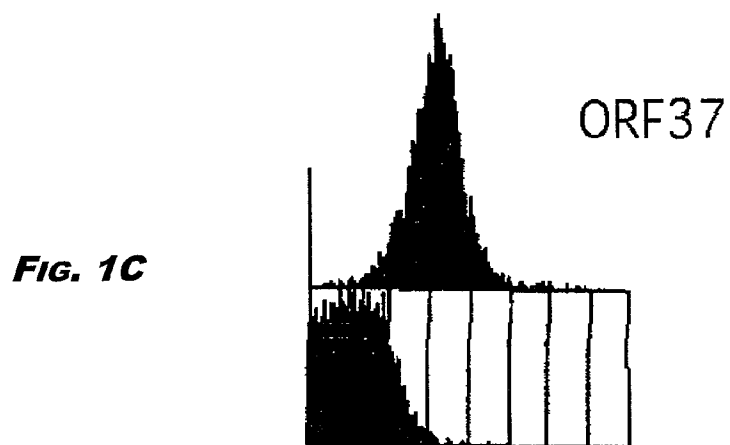
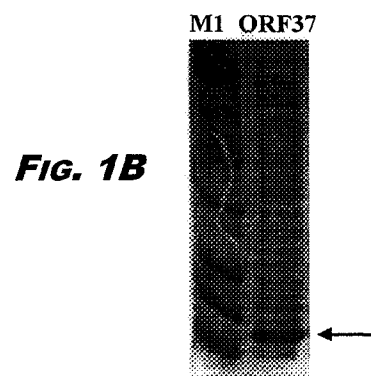
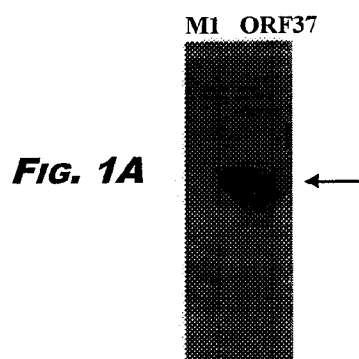
291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,  
331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,  
371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,  
411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,  
5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,  
491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,  
531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,  
571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,  
611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,  
10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,  
691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,  
731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,  
771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,  
811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,  
15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,  
& 891..

10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the  
group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,  
41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,  
20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,  
135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,  
175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,  
215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,  
255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,  
25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,  
335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,  
375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,  
415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,  
455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,  
30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,  
535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,  
575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,  
615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,  
655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,  
35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

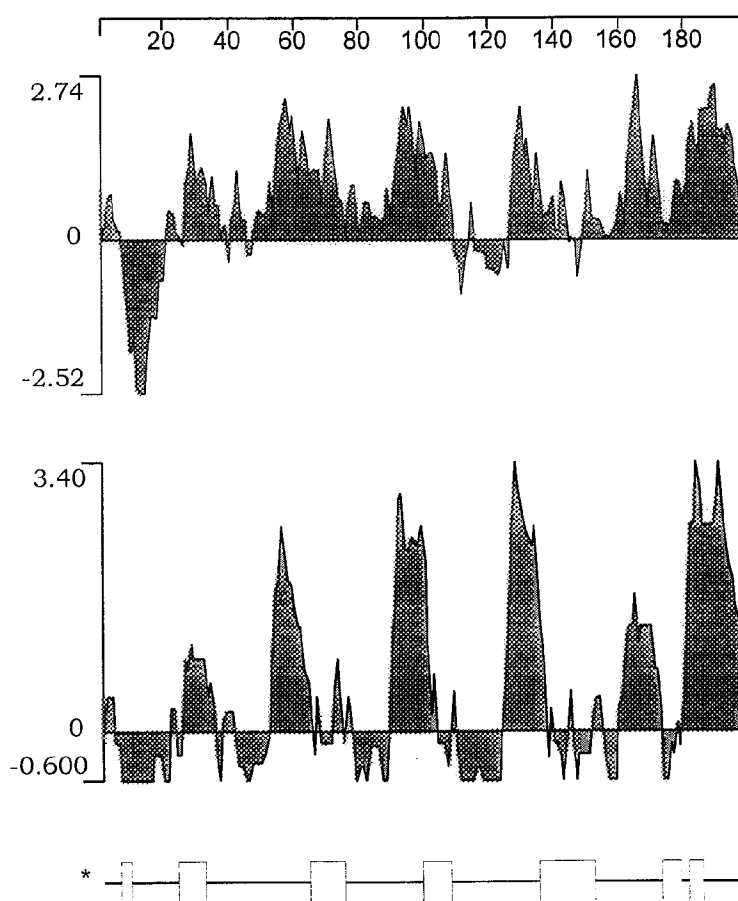
735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any  
10 one of claims 8-12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

1/24

**FIGURE 1**

2/24

**Fig 1E**

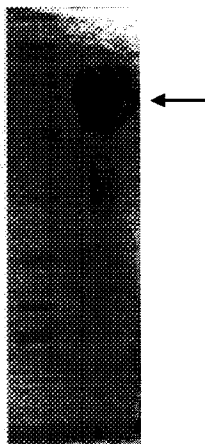


3/24

**FIGURE 2**

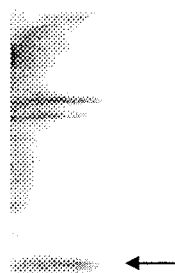
***FIG. 2A***

M1 ORF5

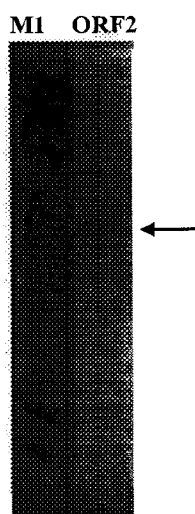
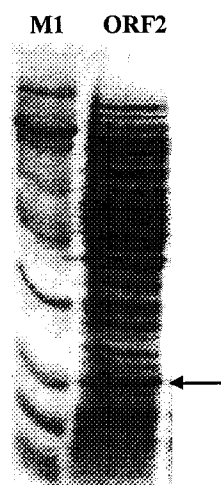
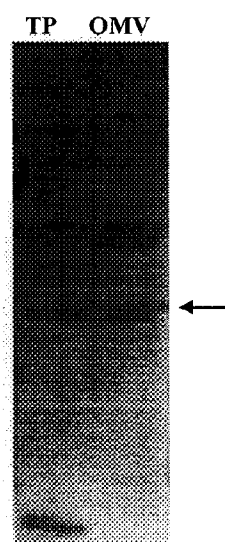
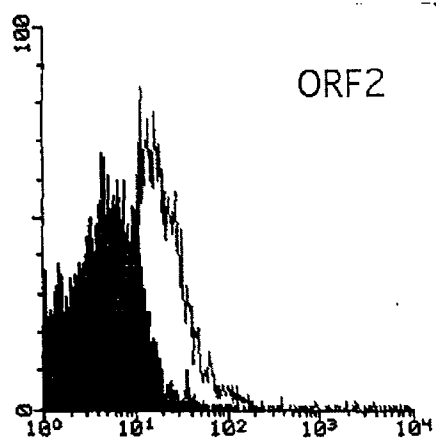


***FIG. 2B***

TP



4/24

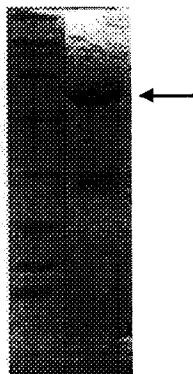
**FIGURE 3****FIG. 3A****FIG. 3B****FIG. 3C****FIG. 3D**

5/24

**FIGURE 4**

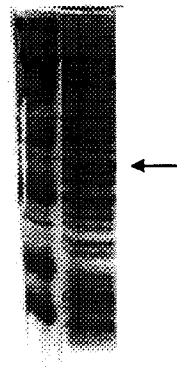
***FIG. 4A***

M1 ORF15



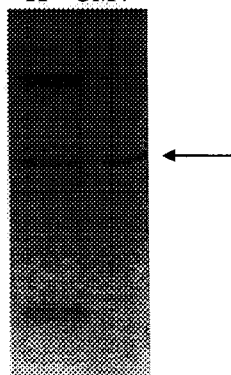
***FIG. 4B***

M2 ORF15

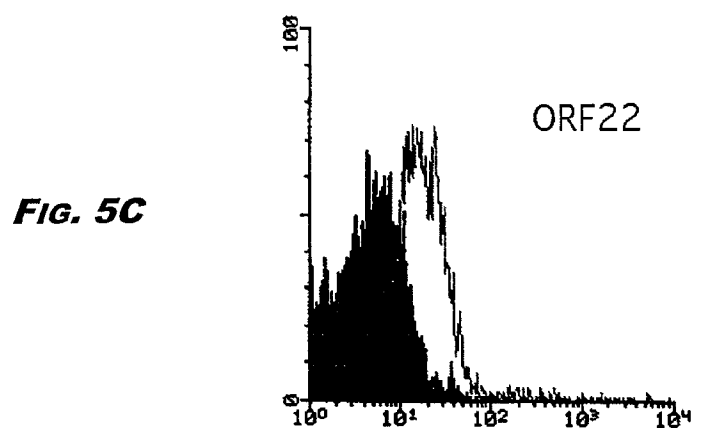
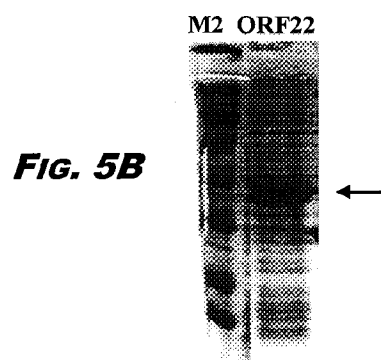
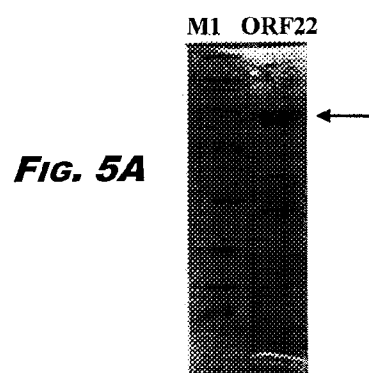


***FIG 4C***

TP OMV



6/24

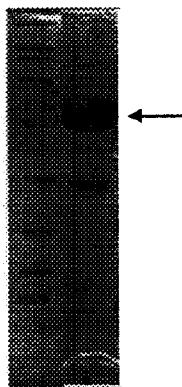
**FIGURE 5**

7/24

**FIGURE 6**

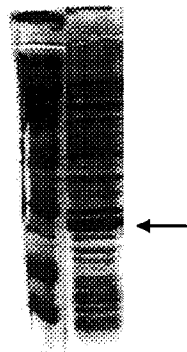
***Fig. 6A***

M1 ORF28



***Fig. 6B***

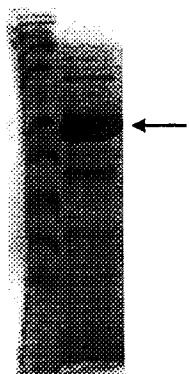
M2 ORF28



**FIGURE 7**

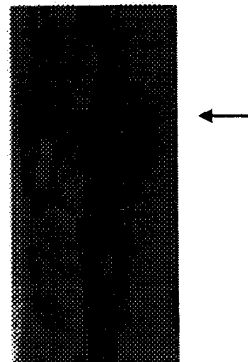
***Fig. 7A***

M1 ORF32

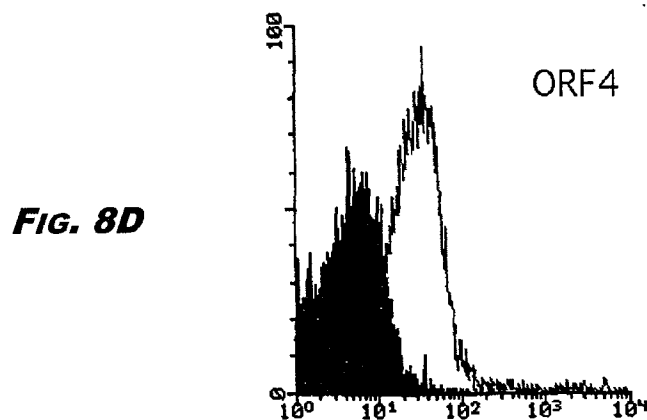
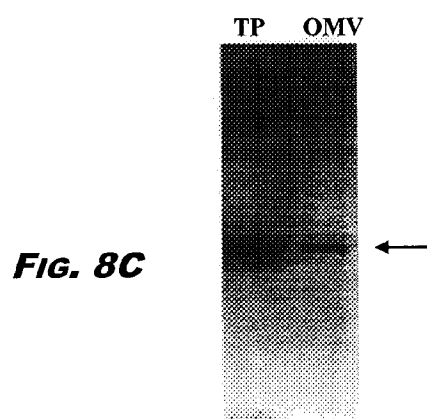
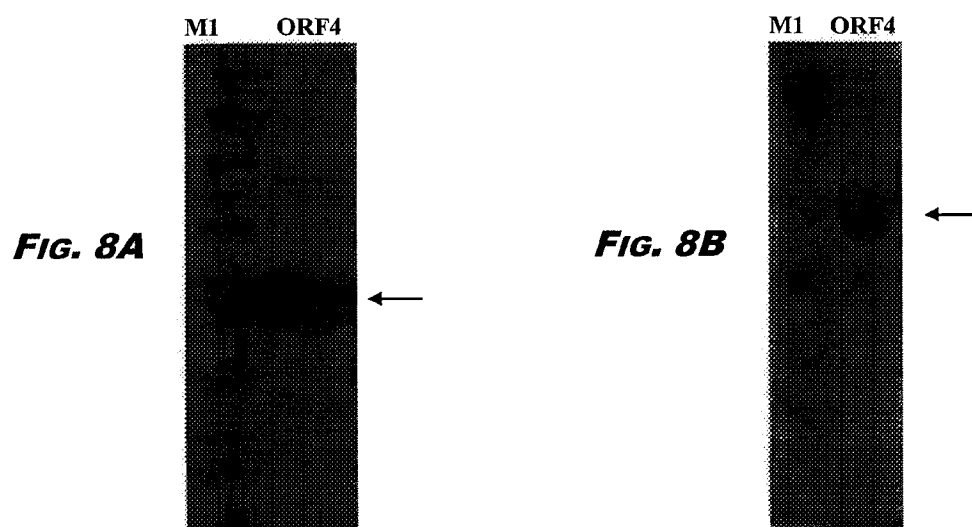


***Fig. 7B***

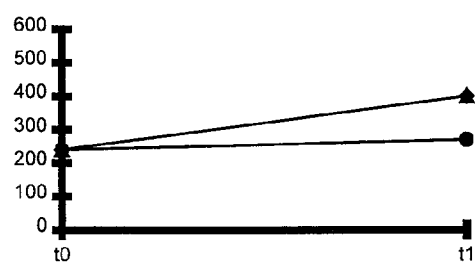
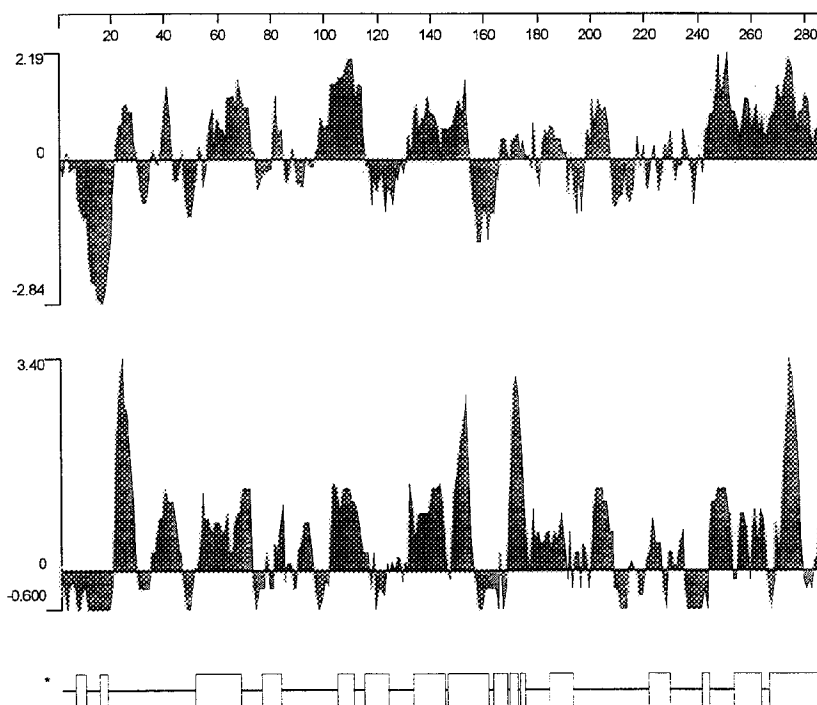
M1 ORF32



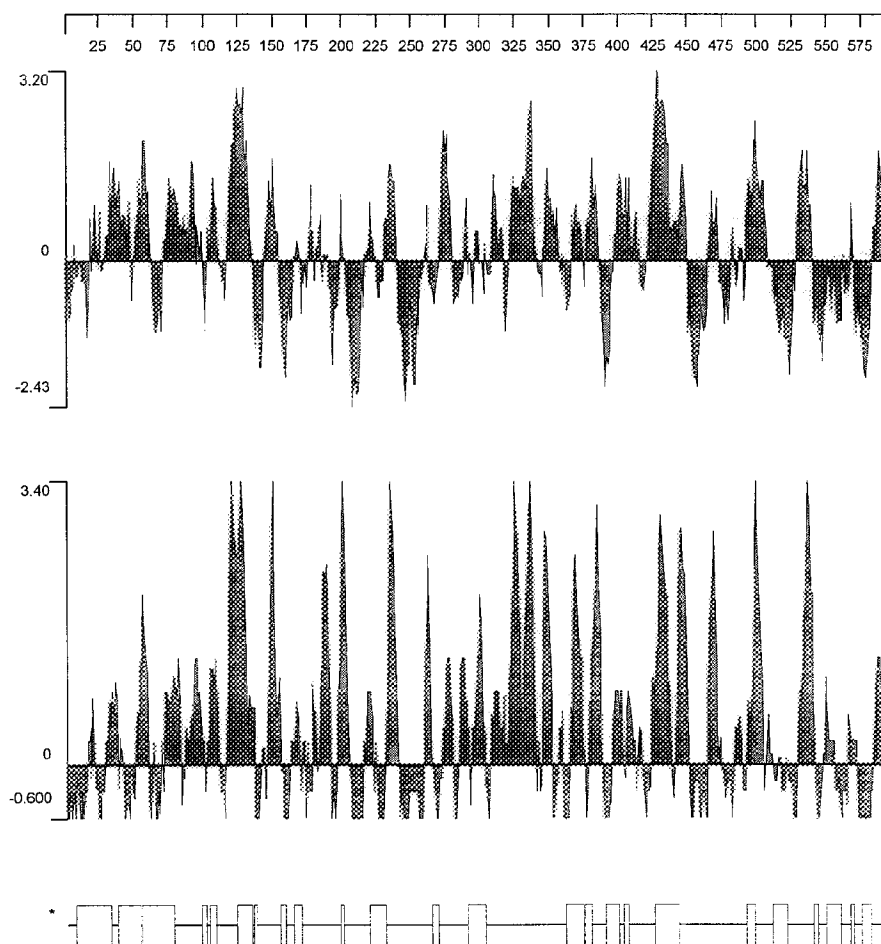
8/24

**FIGURE 8**

9/24

**FIG. 8E****FIG. 8F**

10/24

**FIGURE 9**

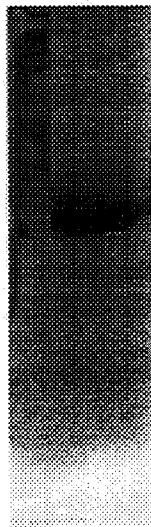


11/24

**FIGURE 10**

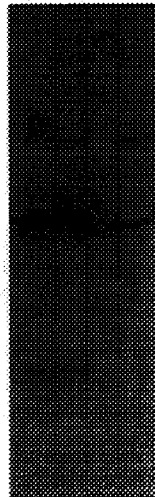
***Fig. 10A***

M1 ORF76

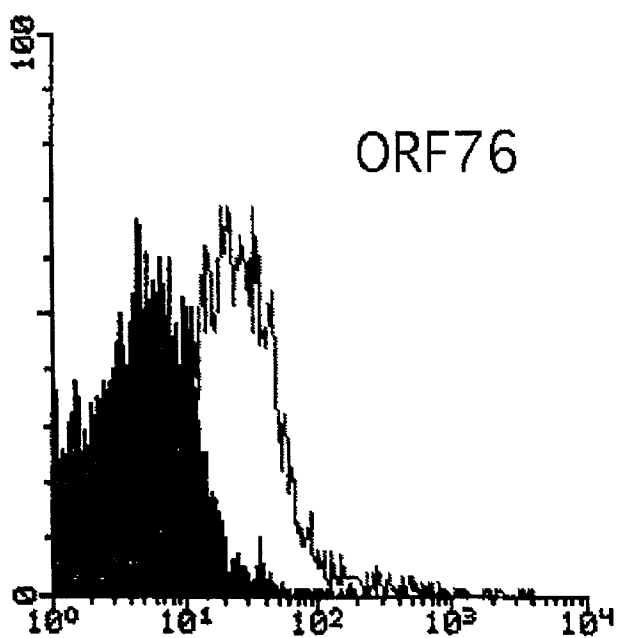


***Fig. 10B***

TP OMV

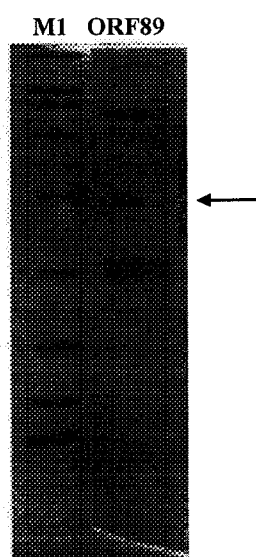


***Fig. 10C***

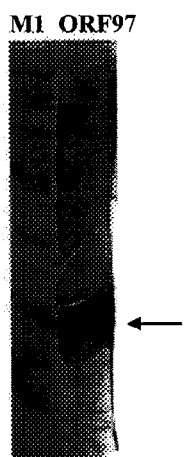
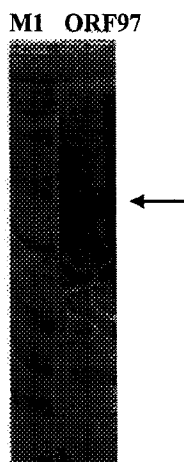
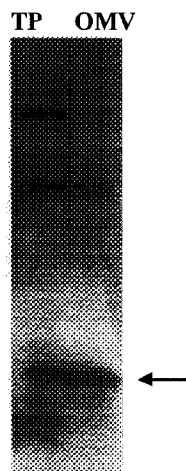
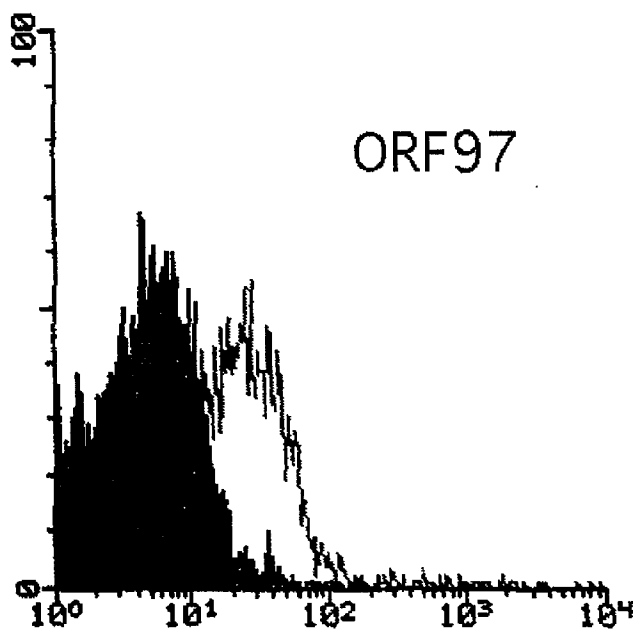


12/24

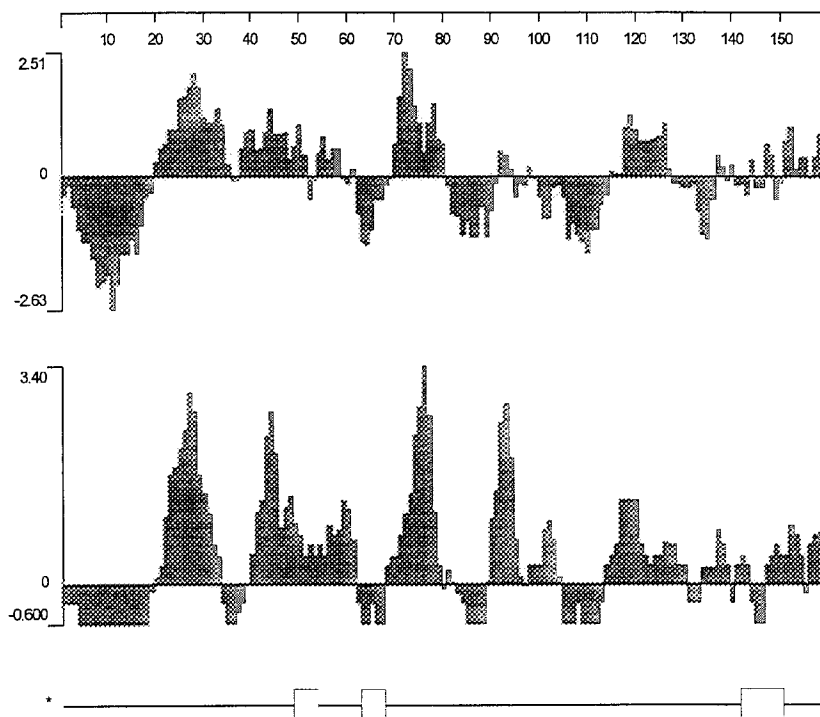
**FIGURE 11**



13/24

**FIGURE 12****FIG. 12A****FIG. 12B****FIG. 12C****FIG. 12D**

14/24

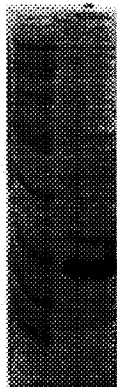
**Fig. 12E**

15/24

**FIGURE 13**

***Fig. 13A***

M1 ORF106

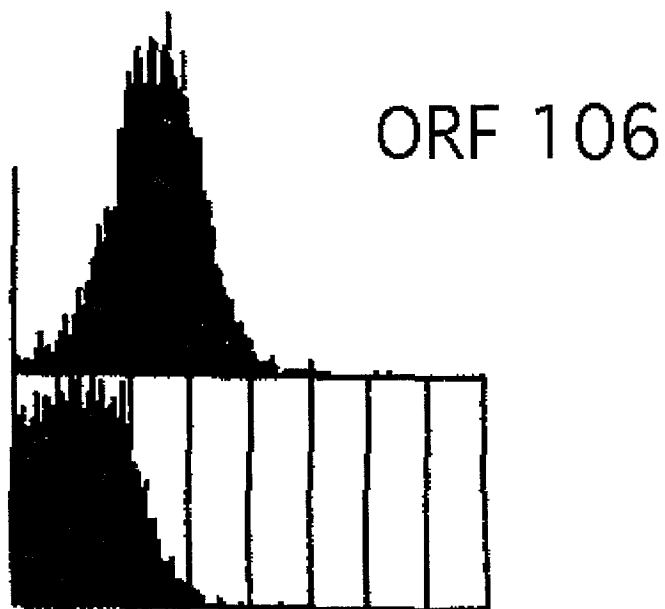


***Fig. 13B***

M2 ORF106

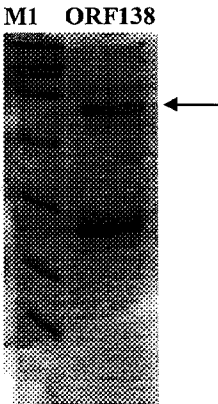


***Fig. 13C***



**FIGURE 14**

***FIG. 14A***



***FIG. 14B***

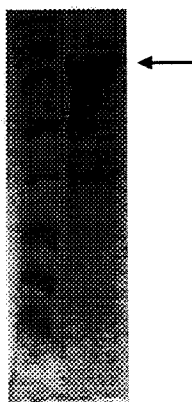


17/24

**FIGURE 15**

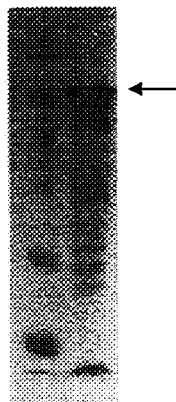
***Fig. 15A***

M1 ORF23



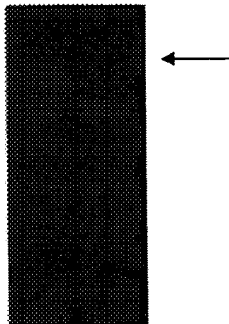
***Fig. 15B***

M2 ORF23

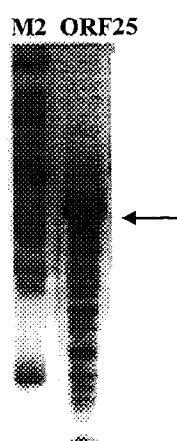
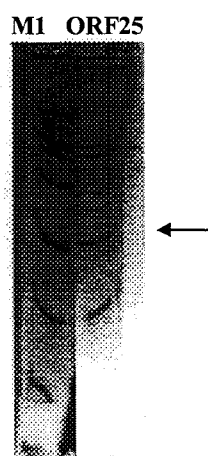
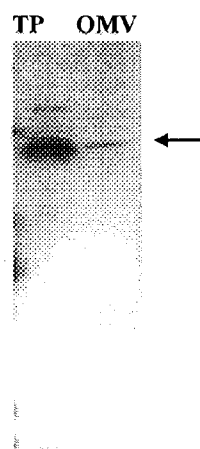
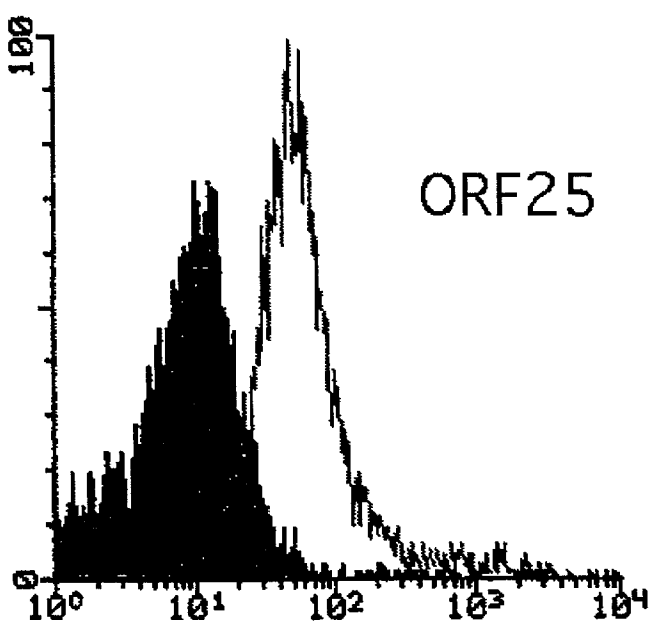


***Fig 15C***

TP OMV

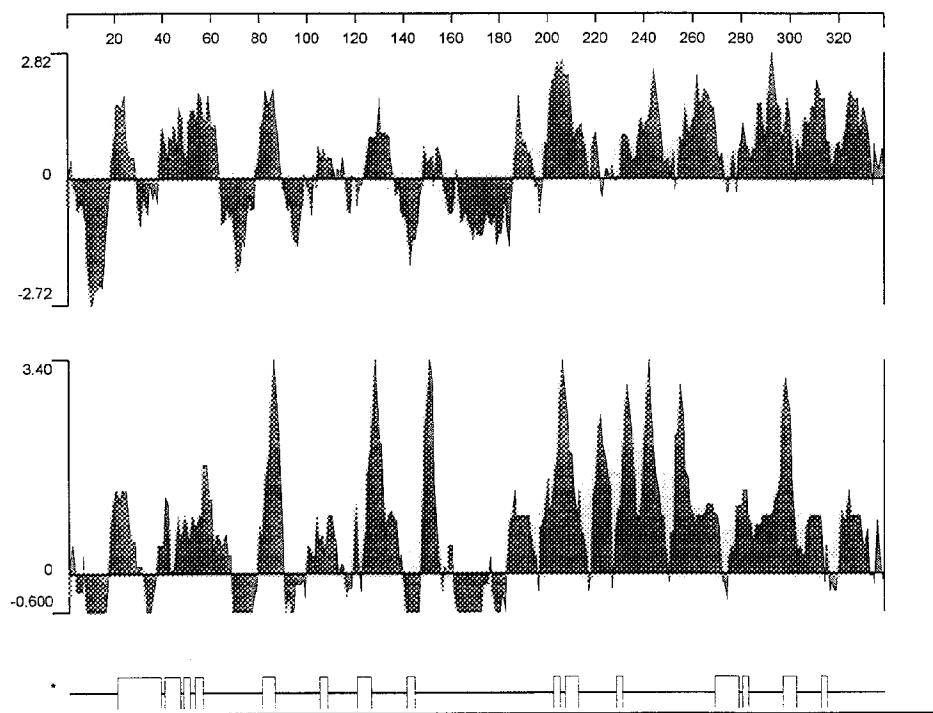


18/24

**FIGURE 16****FIG. 16A****FIG. 16B****FIG. 16C****FIG. 16D**



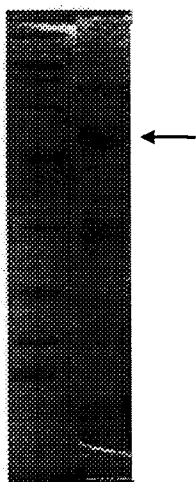
19/24

**FIG. 16E**

**FIGURE 17**

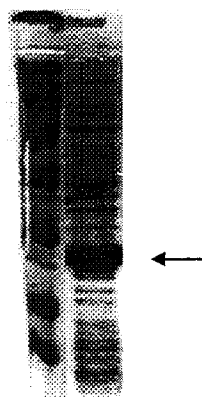
***Fig. 17A***

M1 ORF27

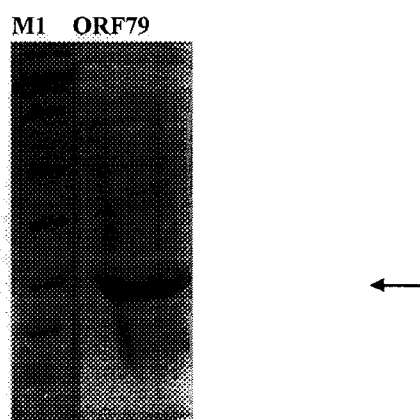
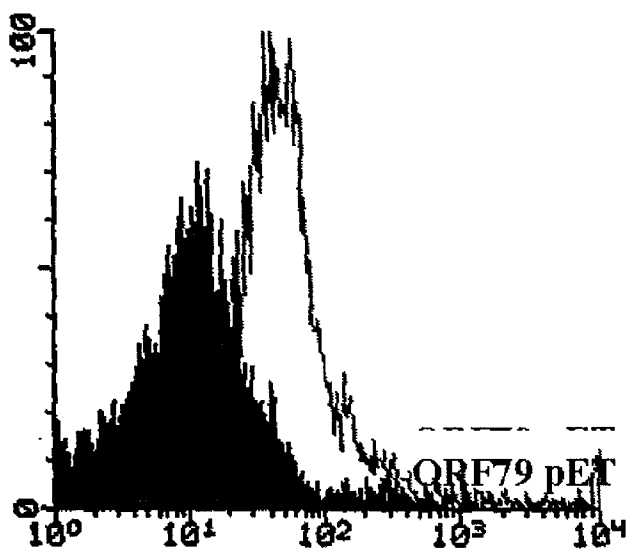


***Fig. 17B***

M2 ORF27



21/24

**FIGURE 18****FIG. 18A****FIG. 18B**

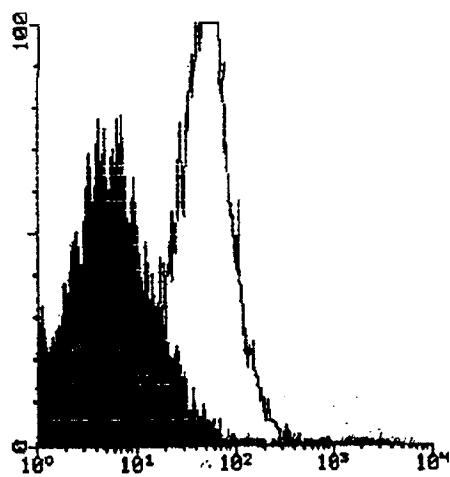
22/24

**FIGURE 19****FIG. 19A**

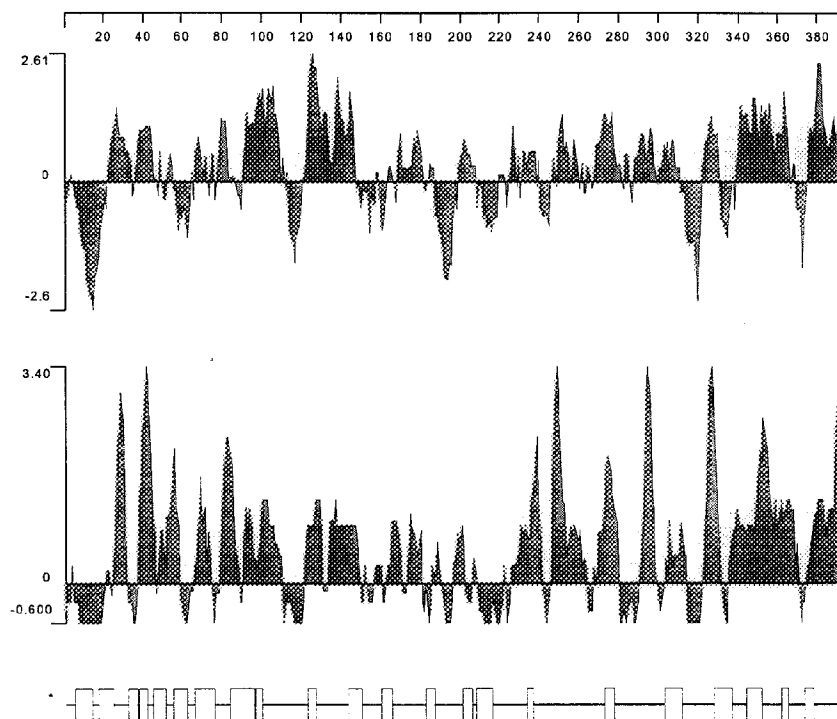
M1 ORF85

**FIG. 19B**

TP OMV

**FIG. 19C**

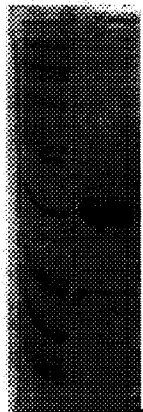
23/24

**Fig 19D**

24/24

**FIGURE 20****Fig. 20A**

M1 ORF132

**Fig. 20B**

M2 ORF132

**Fig. 20C**